

L'expérimentation du revenu de solidarité active entre objectifs scientifiques et politiques

Guillaume Allègre

OFCE, Centre de recherche
en économie de Sciences Po
Département des études

L'expérimentation du revenu de solidarité active conduite dans 34 départements volontaires sur une période allant de novembre 2007 à juin 2009 est sans doute la plus importante expérimentation sociale menée à ce jour en France. L'expérimentation du RSA consiste à évaluer l'effet de plus fortes incitations à l'emploi pour les bénéficiaires du RMI et de l'API en utilisant des groupes tests qui ont accès à un RSA expérimental et des groupes témoins qui n'y ont pas accès. Nous montrons dans une première partie que l'expérimentation d'incitations financières à l'emploi est sujette à de nombreuses difficultés méthodologiques. En particulier, nous montrons qu'il est difficile d'isoler le groupe test pour prévenir la diffusion du traitement expérimental à des personnes non traitées et dans le même temps de construire des groupes expérimentaux permettant un contrôle sans biais. Dans une seconde partie, nous montrons que le protocole mis en place pour l'expérimentation du RSA n'a pas réussi à contourner entièrement ces difficultés. La durée de l'expérimentation n'a pas permis de comparer le caractère pérenne du dispositif par rapport aux dispositifs transitoires en place dans les territoires de contrôle, les biais liés aux effets de mobilisation ont été négligés. De plus, l'éligibilité restreinte au RSA expérimental par rapport au dispositif généralisé n'a pas permis d'évaluer les effets potentiellement négatifs de la familialisation du RSA sur l'incitation à l'emploi des conjoints. Les effets d'éviction sur les travailleurs pauvres non éligibles n'ont pas été mesurés. Enfin, le contexte économique particulièrement défavorable est peu propice aux généralisations. En conclusion, nous émettons l'hypothèse que l'échec relatif de l'expérimentation s'explique par les objectifs en partie contradictoires poursuivis par les acteurs scientifiques et politiques qui l'ont mise en place.

guillaume.allegre@ofce.sciences-po.fr

Mots Clés : Expérimentation. RSA. Minima sociaux.
Incitations à l'emploi. Travailleurs pauvres.

L'expérimentation du revenu de solidarité active (RSA) constitue sans doute la plus importante expérimentation sociale menée à ce jour en France. L'expérience vise à évaluer les effets de plus forts gains financiers au retour à l'emploi des allocataires de minima sociaux. La loi en faveur du travail, de l'emploi et du pouvoir d'achat (dite « loi Tepe ») du 21 août 2007 définit les modalités d'expérimentation du revenu de solidarité active. Selon ce texte de loi, l'objectif du RSA expérimental est « d'assurer l'augmentation des ressources d'une personne bénéficiaire d'un minimum social qui prend ou reprend un travail (...) afin d'atteindre un revenu garanti qui tient compte des revenus d'activité professionnelle et des charges de famille ». La loi d'expérimentation prévoit, pour les départements volontaires, l'autorisation de déroger à certains articles du code de l'action sociale concernant les bénéficiaires du revenu minimum d'insertion (RMI) et de l'allocation de parent isolé (API), dans tout ou partie du territoire départemental. La loi n° 2008-1248 du 1^{er} décembre 2008 institue sur l'ensemble du territoire un revenu de solidarité active au 1^{er} juin 2009 et met ainsi fin à l'expérimentation.

L'expérimentation sociale s'est heurtée en France à des problèmes juridiques et institutionnels. L'article 1^{er} de la Constitution pose le principe du caractère indivisible de la République et celui d'égalité des citoyens devant la loi. Jusqu'en 2003, l'expérimentation se heurtait à ce double obstacle constitutionnel qui empêchait la mise en place d'une législation ou réglementation spécifique sur une portion du territoire (Clément, 2002). Ces contraintes n'ont néanmoins pas totalement empêché la mise en œuvre de politiques sociales expérimentales. Le RMI a notamment fait l'objet d'expérimentations locales préfigurant la loi du 1^{er} décembre 1988. Dès 1985, le département de l'Ille-et-Vilaine introduit un revenu minimum familial garanti, transformé en complément local de ressource (CLR) en 1986. Cependant, ces expériences ont été jusqu'à récemment relativement restreintes. La loi constitutionnelle du 28 mars 2003 relative à l'organisation décentralisée de la République permet d'aller plus loin dans l'expérimentation sociale. L'article 72 de la Constitution prévoit maintenant que « (...) les collectivités territoriales ou leurs groupements peuvent, lorsque selon le cas, la loi ou le règlement l'a prévu, déroger, à titre expérimental pour un objet et une durée limités, aux dispositions législatives ou réglementaires qui régissent l'exercice d'une compétence ». L'expérimentation n'est autorisée que dans l'objectif d'une généralisation : elle ne peut durer plus de 5 ans (avec une prolongation éventuelle de trois ans maximum), et doit faire l'objet d'un rapport d'évaluation ; enfin, la loi doit déterminer à son terme soit son abandon, soit sa généralisation.

L'expérimentation du RSA s'inscrit dans le cadre de la revalorisation du travail et de la lutte contre les trappes à inactivité auxquelles les bénéficiaires de minima sociaux seraient confrontés. L'existence de trappes à inactivité, conséquence de taux marginaux effectifs de prélèvement jugés confiscatoires en cas de reprise d'emploi a été dénoncée à la fin des années 1990 dans plusieurs rapports (CSERC, 1997 ; Bourguignon et Chiappori, 1998). Depuis, un ensemble de mesures visant à réduire

les effets de trappe a été instauré : réforme des aides au logement et de la taxe d'habitation, modification de la décote et du barème de l'impôt sur le revenu, extension puis réformes des possibilités de cumul des minima sociaux et d'une activité rémunérée, création de la prime pour l'emploi (PPE). La PPE vise à accroître le revenu des travailleurs à bas salaires et augmenter les gains financiers à la reprise d'activité. Mais pour de nombreux critiques, la PPE ne poursuit qu'imparfaitement son double objectif de redistribution et d'incitation (Cahuc, 2002).

De même que la PPE, le RSA poursuit un objectif redistributif et un objectif incitatif : combattre la pauvreté des travailleurs à bas salaires et créer un écart entre revenus d'activité et minima sociaux afin d'inciter à la reprise d'emploi. Mais contrairement à la PPE, le RSA est un dispositif familialisé dont le montant tient compte des revenus du foyer et de la situation familiale. Le RSA permet de renforcer les incitations à l'emploi là où elles sont actuellement faibles, c'est-à-dire à la reprise d'emploi à temps partiel pour les célibataires et au premier emploi à temps plein pour les couples bénéficiaires de minima sociaux (Allègre, 2008). Cependant, le RSA ne peut avoir un effet sur l'emploi que sous certaines conditions. L'expérimentation du RSA a notamment pour objet de vérifier si ces conditions sont actuellement réunies en France.

Le RSA, comme son expérimentation, ont des précédents à l'étranger. Le RSA a été proposé pour la première fois dans un rapport de la commission « Familles, vulnérabilité, pauvreté » (Hirsch, 2005). Il est similaire à l'Allocation compensatrice de revenus (ACR) proposée en France par Godino (1999) et s'inspire de l'*Earned Income Tax Credit* (EITC), mis en œuvre aux États-Unis en 1975, et du *Working Family Tax Credit* (WFTC), en œuvre au Royaume-Uni entre 1999 et 2003¹. L'expérimentation des incitations financières à l'emploi pour les personnes peu qualifiées s'est développée aux États-Unis dans les années 1960 et 1970. Plus récemment, le projet d'autosuffisance mené à partir de 1992 au Canada a fait l'objet d'une vaste littérature (e.g. Allègre, 2008).

L'expérimentation du RSA a été menée dans 34 départements volontaires sur une période allant de novembre 2007 à juin 2009. Conformément à la loi, un comité composé de représentants des conseils généraux, de représentants de l'État et des organismes payeurs et de personnalités qualifiées a été chargé d'évaluer les expérimentations. Un rapport intermédiaire est paru en septembre 2008 et le rapport final en mai 2009 (Comité d'évaluation des expérimentations, 2008 et 2009). L'objectif de notre étude est d'analyser le dispositif expérimental mis en place, de montrer comment il tente de contourner les difficultés liées à l'expérimentation sociale des incitations financières et de tirer les leçons de l'expérimentation, à la fois sur les connaissances économiques concernant l'impact des gains financiers sur l'offre de travail et sur la pratique de l'expérimentation dans ce domaine.

1. Le WFTC a été remplacé par un instrument non familialisé, mais conjugalisé en 2003. Pour une revue de littérature concernant les évaluations du WFTC, voir Blundell et Walker (2001).

Dans une première section nous présentons l'intérêt de l'expérimentation contrôlée dans le cadre de l'évaluation des compléments de revenus pour travailleurs et nous discutons les questions méthodologiques. Dans une deuxième section nous analysons à cette lumière le protocole mis en place dans le cadre de l'expérimentation du RSA et nous commentons les résultats de l'expérimentation. Enfin, nous concluons sur l'apport des expérimentations sociales de grandes ampleurs dans le cadre de l'évaluation des politiques publiques.

1. L'expérimentation contrôlée d'incitations financières à l'emploi : intérêt et questions méthodologiques

1.1. Une solution au biais de sélection

Quel serait l'effet sur l'emploi et le chômage d'une augmentation des gains financiers à l'emploi ? Si cette question paraît simple, il est en fait difficile d'évaluer l'élasticité d'offre de travail sur des données non expérimentales. On pourrait vouloir comparer la situation des personnes ayant des forts gains financiers à celle ayant de faibles gains mais nous ne pouvons connaître avec certitude le salaire auquel peuvent prétendre les personnes n'ayant pas d'emploi. Sans ce salaire potentiel, nous ne connaissons pas les gains à la reprise d'emploi des personnes ne travaillant pas. La méthode la plus simple consiste à estimer une équation de salaire sur les personnes en emploi afin de déduire le salaire potentiel des personnes qui ne travaillent pas. Mais la population des individus employés et celle des individus inemployés ont probablement des caractéristiques inobservables très différentes qui peuvent avoir une influence sur leur statut d'emploi. Si l'on ne peut résoudre ce problème de biais de sélection, on ne peut pas séparer de manière certaine l'impact des incitations financières des autres facteurs de non-emploi.

L'expérimentation est, en théorie, une solution simple au problème du biais de sélection. Comme le soulignent Heckman et Smith (1995), le problème fondamental de l'évaluation des politiques publiques est l'impossibilité d'observer une personne donnée dans l'état où la politique publique à évaluer est appliquée et au même moment dans celui où elle n'est pas appliquée. Si cela était possible, il suffirait de comparer les deux états pour connaître l'effet de la politique publique selon le modèle d'inférence causale de Rubin (1974). Puisqu'il est impossible d'observer un individu à un moment donné dans deux états différents, le principe des expériences contrôlées est de comparer deux groupes identiques d'individus, un groupe étant soumis à la politique publique à évaluer, l'autre n'y étant pas soumis. Ces expériences s'inspirent des expérimentations médicales dont elles utilisent le vocabulaire : la politique à évaluer est appelée « traitement », le groupe qui y est soumis est le « groupe de traitement », celui qui n'y est pas soumis est le « groupe de contrôle », le groupe de traitement et le groupe de contrôle forment l'échantillon et l'expérimentation est une « expérience contrôlée ». Selon le modèle causal de Rubin, l'effet causal d'un traitement est la différence entre ce qui arrive en étant

exposé à un traitement et ce qui serait arrivé en n'y étant pas exposé (notion de contrefactuel). Selon ce raisonnement, on conclut dans les expérimentations que le traitement est la cause de la différence entre les deux groupes sous deux conditions : (1) les deux groupes doivent être effectivement identiques ; (2) le traitement doit être la seule circonstance qui diffère entre les deux groupes et le traitement ne doit pas affecter le groupe de contrôle (absence de diffusion du traitement). Si ces deux conditions ne sont pas remplies, la validité interne de l'expérience n'est pas respectée et l'estimation de l'impact du traitement sera biaisée.

Afin de s'assurer que le groupe de contrôle et le groupe de traitement partagent les mêmes caractéristiques, il est usuel d'affecter les individus participant à l'expérimentation aux groupes de traitement et de contrôle de manière aléatoire. Cette méthode est utilisée depuis les travaux de Fisher dans les années 1920. Elle part du principe qu'il est impossible de créer deux groupes – d'individus ou d'objets – parfaitement identiques. Il est donc nécessaire de faire en sorte que les différences entre les deux groupes, plus ou moins grandes, ne biaisent pas l'estimation de l'impact du traitement. Selon Fisher (1935), la méthode du tirage aléatoire est la plus satisfaisante à cet égard. Les différences de résultats entre deux groupes tirés au sort seront la conséquence (1) de différences aléatoires, (2) d'effets intervenant au hasard, indépendants du traitement et (3) de différences de réactions au traitement à tester. L'expérimentation consiste à tester l'hypothèse nulle d'absence de différences de réaction au traitement. Les expériences contrôlées sont représentatives des méthodes déductives : si les hypothèses concernant les conditions de validité sont justes, alors leurs conclusions doivent être justes. En cela, elles peuvent être qualifiées de concluantes (Cartwright, 2007).

Outre leur validité scientifique et leur caractère concluant, un avantage souvent cité des expériences contrôlées est la facilité d'interprétation de leurs résultats. Il suffit en effet de comparer le comportement des groupes expérimentaux avant et après le traitement par une méthode de double différence pour calculer l'impact moyen du traitement. Des techniques économétriques peuvent être utilisées pour introduire d'autres contrôles mais le principe reste le même : il s'agit simplement de calculer un écart entre le groupe test et le groupe témoin. Le protocole est ainsi plus facilement compréhensible que ne peuvent l'être les études sur données non expérimentales. La simplicité de l'analyse des résultats permettrait de faciliter la discussion au sein de la communauté scientifique ainsi que la communication entre celle-ci et les personnes en charge des politiques publiques. Autre avantage de l'expérimentation, elle permet d'évaluer des programmes qui n'ont pas été mis en place auparavant, soit parce que totalement nouveaux, soit d'une ampleur sans précédent. Dans ce cas, l'extrapolation de données non expérimentales peut engendrer des estimations biaisées.

1.2. Questions éthiques

L'expérimentation sociale pose un certain nombre de problèmes éthiques et juridiques puisque, dans son principe, certaines personnes éligibles se voient refuser

de manière aléatoire l'accès à un traitement potentiellement bénéfique. À l'inverse, à quelles conditions peut-on soumettre à un groupe d'individus un traitement qui leur est potentiellement nuisible ? Dans les deux cas, la question du coût social et politique de l'expérimentation se pose. Dans le cadre de l'évaluation des incitations financières, on ne peut réellement imaginer un traitement qui ne serait pas, au moins potentiellement, bénéfique : les objections face à une expérimentation d'une baisse des prestations sociales seraient probablement trop importantes pour que l'expérience puisse être réellement conduite. En expérimentation sociale, le traitement est ainsi le plus souvent bénéfique de façon non ambiguë. Dans le cadre des incitations financières à l'emploi, on refuse un avantage monétaire à des personnes du groupe de contrôle alors qu'il est reçu par des personnes du groupe de traitement ayant le même comportement. Le problème éthique est d'autant plus accentué que les transferts sont élevés et que l'échantillon concerné par l'expérience est important. Une solution proposée par Burtless et Orr (1986) est de compenser les membres du groupe de contrôle mais ceci est difficile à réaliser lorsque le traitement à évaluer est une prime monétaire. Une autre solution est d'informer les individus concernés et d'obtenir leur consentement : lorsqu'il s'agit d'un avantage monétaire, il est peu probable que des individus refusent de participer à l'expérience.

1.3. Menaces sur la validité interne

La validité interne d'une expérience est sa capacité à produire une estimation non biaisée de l'effet du traitement sur l'échantillon participant à l'expérimentation. Elle est menacée (1) lorsqu'il existe des différences de caractéristiques entre les groupes de contrôle et de traitement qui ne sont pas aléatoires et qui affectent la réaction au traitement ; (2) lorsque le traitement à évaluer n'est pas le seul paramètre qui diffère entre les deux groupes et (3) lorsque le groupe de contrôle est affecté par le traitement. La validité de l'expérience requiert donc l'absence de diffusion du traitement dite *stable unit treatment value assumption* (SUTVA) dans le modèle causal de Rubin (1974). Cartwright (2007) souligne que pour retirer le bénéfice principal d'une expérience contrôlée – son caractère concluant – toutes les hypothèses doivent être réunies. En effet, le raisonnement déductif est soumis au principe du maillon faible : la robustesse de la conclusion ne peut être supérieure à celle de l'hypothèse la moins robuste.

Les différences de caractéristiques entre les deux groupes expérimentaux peuvent avoir plusieurs sources. Les individus des deux groupes peuvent abandonner de manière différentielle l'expérience, et ce à plusieurs stades. Ils peuvent décider de refuser de participer à l'expérience (biais de participation) ou de l'abandonner après quelques mois (biais d'attrition). Lorsque les résultats sont évalués par enquête, les enquêteurs peuvent également avoir des difficultés à collecter les données, soit parce qu'ils ne retrouvent pas les individus concernés, soit parce que ceux-ci refusent de répondre (biais de non-réponse). Si le traitement offre un bénéfice élevé à certains individus, comme c'est le cas avec les transferts monétaires, la non-réponse et

l'attrition risquent d'être fortement corrélées à la réaction au traitement : elles ont de fortes chances d'être plus élevées dans le groupe de contrôle et chez les individus non potentiellement bénéficiaires du groupe de traitement (ils ont une probabilité plus élevée de déménager et de ne pas répondre aux enquêteurs si la réponse à l'enquête demande un effort). Ce type de biais de sélection peut alors être comparable à celui que l'on trouve dans les données non expérimentales (Burtless et Orr, 1986).

Afin d'obtenir des estimations non biaisées, le protocole expérimental doit être mis en œuvre de telle sorte que le traitement à évaluer soit la seule différence parmi les circonstances auxquelles sont confrontés les deux groupes expérimentaux. Il faut en premier lieu que le traitement soit effectivement donné à toutes les personnes du groupe de traitement et refusé à celles du groupe de contrôle. Il ne faut pas que le groupe de contrôle ait accès à un substitut ou que le groupe de traitement perde l'accès à des programmes complémentaires. Dans le cadre d'un avantage monétaire, il faut s'assurer que celui-ci constitue bien la différence de revenu, à circonstances égales, entre un membre du groupe de contrôle et un membre du groupe de traitement. Si l'expérimentation nécessite le travail de services sociaux, il faut également que les travailleurs sociaux agissent avec les deux groupes de manière équivalente. Par exemple, si l'expérience nécessite une campagne d'information, celle-ci devrait aussi concerner le groupe de contrôle. L'expérimentation sociale peut aussi faire l'objet d'enjeux politiques : si les travailleurs sociaux veulent voir l'expérience réussir – ou au contraire échouer – ils auront tendance, de manière consciente ou inconsciente, à traiter différemment les membres du groupe de traitement et ceux du groupe de contrôle. De plus, si les acteurs sociaux refusent le caractère aléatoire de l'accès au traitement, ils peuvent orienter les membres du groupe de contrôle vers des programmes de substitution.

Les participants à l'expérience peuvent changer de comportement parce qu'ils se savent l'objet d'une expérimentation. On parle d'effets de mobilisation ou « effet Hawthorne », d'après le nom d'une usine près de Chicago où Elton Mayo mena une série d'expérimentations entre 1924 et 1932 sur la productivité des ouvriers selon les conditions de travail. Les résultats montrèrent une hausse de la productivité quelles que soient les modifications des conditions de travail (hausse et baisse de la luminosité, du nombre de pauses, etc.), et même lorsqu'étaient remises les conditions initiales de travail. L'effet Hawthorne biaise les résultats d'une expérience contrôlée si les groupes d'expérimentation réagissent différemment, par exemple si les membres du groupe de contrôle sont découragés du fait de se voir refuser un traitement potentiellement bénéfique ou si les membres du groupe de traitement sont motivés non pas par le traitement lui-même mais par le fait d'être traité. L'effet Hawthorne se rapproche de l'effet placebo en médecine. Afin de réduire l'influence que peut avoir la connaissance des groupes expérimentaux sur le patient et sur l'équipe médicale, les expériences biomédicales contrôlées sont, lorsque c'est possible, effectuées en double aveugle : ni les patients, ni l'équipe médicale ne savent qui utilise les produits actifs et qui utilise les placebos.

L'expérience en double aveugle est évidemment plus difficile en expérimentation sociale. Elle est impossible dans le cadre de l'évaluation des incitations financières : dans ce cas, les membres du groupe de traitement doivent nécessairement être informés de l'existence de nouvelles incitations et donc avertis de leur participation à l'expérience et de leur appartenance au groupe de traitement. Card et Robbins (1996) soulignent un dilemme auquel sont confrontés les évaluateurs dans ce type d'expérimentation : d'une part, il est important que les individus connaissent en détail le programme et les bénéfices potentiels d'un retour à l'emploi mais d'autre part, il faut minimiser les effets Hawthorne liés à un traitement spécial des participants. Une solution à ce dilemme est d'informer de manière équivalente les personnes composant le groupe de contrôle de leurs droits existants.

L'hypothèse d'absence de diffusion du traitement est une hypothèse déterminante quant à la validité des expérimentations contrôlées. Dans le cas contraire, la différence de réaction entre le groupe de contrôle et le groupe de traitement ne sera pas une mesure sans biais de l'effet du traitement. En effet, l'hypothèse implicite de l'expérimentation contrôlée est que le groupe de contrôle se comporte comme le groupe de traitement se comporterait en l'absence de traitement : le groupe de contrôle est le contrefactuel du groupe de traitement non traité. Or, ce n'est pas le cas lorsque le groupe de contrôle est affecté indirectement par le traitement. L'exemple médical du vaccin est parlant. Si l'on tire au sort dans une population donnée un groupe de traitement que l'on vaccine contre une maladie contagieuse, et un groupe de contrôle que l'on ne vaccine pas, le fait de vacciner le groupe de traitement va réduire la contagion et donc également le nombre de personnes atteintes dans le groupe de contrôle (on parle de diffusion du traitement au groupe de contrôle ou « *spill-over* » lorsqu'il bénéficie indirectement du traitement). La différence entre le nombre de personnes atteintes entre le groupe de contrôle et le groupe de traitement sous-estimera l'effet bénéfique du vaccin. Dans ce cas, le groupe de contrôle n'est plus un bon contrefactuel. La différence entre les deux groupes n'est pas l'effet du traitement sur le premier groupe mais l'effet du traitement sur le premier groupe moins l'effet du traitement sur le deuxième groupe. Si l'on veut isoler l'effet du traitement, il faut isoler les deux groupes. Dans le cadre d'une expérimentation sur le terrain, on ne peut tirer au sort des individus lorsqu'il existe des effets de diffusion. Dans le domaine de l'emploi, il peut y avoir diffusion du traitement si le fait que mes voisins ou connaissances travaillent augmente ma probabilité de travailler ou de chercher du travail. Sur le marché du travail, les effets d'éviction sont peut-être plus répandus. Il y a éviction lorsque le bénéfice du traitement se fait au détriment des personnes non traitées. S'il existe du chômage involontaire, des instruments ciblés sur certaines catégories de publics peuvent avoir des effets importants sur l'emploi du groupe traité et avoir des effets réduits (voire nuls) au niveau global. Selon Burtless et Orr (1986), ces effets d'éviction peuvent être un problème important lorsque l'on veut évaluer l'effet de l'assistance à la recherche d'emploi ou de programmes de formation sur la reprise d'emploi de personnes peu qualifiées : les auteurs considèrent que ce type de

traitement leur confère un avantage significatif aux dépens des personnes n'ayant pas accès à ce type de programmes.

De manière plus générale, un problème fondamental des expériences avec tirage aléatoire est qu'elles ne capturent pas tous les effets d'équilibre général. Sur le marché du travail, le niveau de l'emploi et des salaires est le résultat d'un équilibre entre l'offre et la demande de travail. Dans ce cadre, expérimenter une mesure sur une partie des travailleurs (ce qui est nécessaire lorsqu'on veut former un groupe de contrôle par tirage aléatoire) n'aura pas les mêmes effets qu'une généralisation à tous les travailleurs. Par exemple, si la mesure que l'on expérimente augmente l'offre de travail du groupe traité, en faisant l'hypothèse d'une demande élastique, l'augmentation de l'offre de travail aura pour conséquence une baisse du salaire d'équilibre, ce qui, dans un modèle d'offre de travail, aura pour effet une réduction de l'emploi dans le groupe de contrôle. L'écart du taux d'emploi entre le groupe de traitement et le groupe de contrôle sera alors une estimation biaisée de l'effet du traitement. Dans le cas extrême d'une demande totalement inélastique, l'augmentation de l'emploi dans le groupe de traitement sera compensée intégralement par une diminution de l'emploi dans le groupe de contrôle. La littérature montre que ces effets d'équilibre général peuvent être importants sur le marché du travail (Lise *et alii*, 2005). Dans une évaluation des effets d'une subvention aux salaires, Blundell *et alii* (2003) construisent un modèle en équilibre général d'offre de travail et d'accumulation de capital humain. Ils comparent alors les effets de la subvention aux salaires en équilibre général à ceux qui seraient observés en équilibre partiel lors d'une expérimentation. Les auteurs montrent que les effets en termes de chômage qui seraient observés en équilibre partiel sont inverses aux effets en équilibre général. Van der Linden (2005) évalue les politiques d'activation sur le marché du travail dans un modèle en équilibre général de *Job search* avec négociations salariales par les salariés en place au nom de tous les travailleurs. L'auteur montre alors que les politiques d'activation ont un effet positif sur le taux d'emploi des participants aux programmes mais détériore la situation des non-participants à travers la formation des salaires et des ajustements dans leur recherche d'emploi. Il conclue que l'impact apparent d'un tel programme dans une évaluation expérimentale différerait grandement de son impact en équilibre général.

1.4. Menaces sur la validité externe et reproductibilité

La validité externe d'une expérience est sa capacité à être généralisée à l'ensemble de la population concernée. Pour cela, (1) l'échantillon doit être représentatif de la population, (2) la réaction au traitement et au non-traitement doit être indépendante du nombre de personnes auxquelles il a été donné et (3) la réaction au traitement ne doit pas dépendre du fait que celui-ci ne peut être administré que de façon temporaire lors d'une expérimentation. De plus, pour que l'expérimentation ait un sens, (4) le traitement doit pouvoir être reproduit à une plus grande échelle et (5) l'effet doit être stable dans le temps.

Pour que l'estimation de l'effet du traitement ne soit pas biaisée, le groupe participant à l'expérimentation doit être représentatif de la population. En fait, il suffit qu'il n'y ait pas de corrélation entre la probabilité de faire partie de l'échantillon et la réaction au traitement. Sinon, il existe un « biais d'échantillonnage ». Dans les expériences contrôlées aléatoires, la probabilité de faire partie de l'échantillon ne doit pas être affectée par le fait que le traitement soit conditionné par un tirage au sort. Pour assurer la représentativité des personnes participant à l'expérience, la méthode la plus satisfaisante consiste à effectuer un double tirage au sort, de l'échantillon dans la population concernée puis des groupes d'expérimentation dans l'échantillon.

La réaction au traitement et au non-traitement doit être indépendante du nombre de personnes auxquelles il a été donné. Cette condition n'est pas respectée lorsque le traitement a des effets sur l'équilibre d'un marché. En cela, ces effets d'équilibre menacent à la fois la validité interne et externe de l'expérimentation. Un effet d'équilibre pouvant affecter la validité externe d'une expérimentation est l'effet d'auto-sélection. On parle d'auto-sélection lorsque des individus modifient leurs comportements pour devenir éligible à un programme. Prenons un exemple théorique d'inspiration néo-classique : un complément de revenu pour travailleurs à bas salaires financé par une imposition sur les plus hauts salaires peut avoir pour effet de réduire le rendement monétaire de l'éducation ; et donc, si certains individus réagissent, de réduire le niveau de qualification moyen et d'augmenter le nombre de personnes potentiellement éligibles aux compléments de revenu. Ce type de phénomène sera sous-estimé dans une expérimentation où l'éligibilité au programme ne peut pas être anticipée. Il est possible mais difficile de tester l'auto-sélection par des protocoles expérimentaux car l'auto-sélection sera nettement plus importante lorsqu'un programme est jugé stable.

Les expérimentations ont par construction une durée limitée. L'évaluation de l'effet d'un programme pérenne sera biaisée si les agents réagissent différemment à un traitement temporaire. Ce biais de durée limitée peut être important lorsque le traitement est un transfert monétaire. Prenons l'exemple d'un transfert monétaire qui rend le travail plus rémunérateur. Le caractère temporaire d'un tel transfert aura deux effets sur le changement de l'offre de travail des individus concernés. L'effet de substitution sera plus marqué du fait des arbitrages inter-temporels : si par exemple il est plus rémunérateur de travailler aujourd'hui que demain, certaines personnes peuvent repousser une sortie temporaire du marché du travail (par exemple pour cause de congé parental). D'autre part, dans un cadre de cycle de vie, lorsque l'incitation financière est temporaire, le revenu espéré pour l'ensemble des années à venir (« revenu permanent ») est plus faible que dans le cadre d'un transfert monétaire pérenne. Si le loisir est un bien normal, la diminution de l'offre de travail due à cet « effet revenu » sera donc plus faible que dans le cadre d'un programme pérenne. De part ces deux effets, l'offre de travail sera plus élevée lorsque une subvention à l'emploi est temporaire que lorsqu'elle est pérenne. Une expérimentation risque alors de surestimer une éventuelle augmentation de l'offre

de travail liée à des suppléments de revenu d'activité. Dans l'exemple inverse d'un transfert monétaire qui rend l'inactivité plus rémunératrice, le caractère temporaire du transfert aura deux effets de sens opposé. De part l'effet de substitution, l'expérimentation surestimera la réduction de l'offre de travail : s'il est plus rémunérateur de ne pas travailler aujourd'hui que demain, certaines personnes anticiperont une sortie temporaire du marché du travail. Par contre, de part l'effet de revenu permanent, la diminution de l'offre de travail sera plus faible dans une expérimentation. L'introduction temporaire d'un revenu minimum garanti peut donc surestimer ou sous-estimer l'effet permanent d'un tel programme. Metcalf (1974) évalue ce biais de durée limitée dans une expérience au New Jersey en utilisant des informations sur les comportements de consommation des foyers. Une autre solution pour évaluer ce biais consiste à expérimenter plusieurs durées du traitement (Burtless et Greenberg, 1982 ; Robbins, 1983).

La durée limitée d'une expérimentation empêche également les effets d'apprentissage : la réaction à court terme à un traitement peut être différente de celle d'un équilibre de moyen ou de long terme car les personnes concernées peuvent apprendre de nouvelles informations au fur et à mesure du programme. Par exemple, ils peuvent dans un premier temps sous-évaluer ou surévaluer le gain monétaire à reprendre un emploi. S'ils surévaluent le gain à reprendre un emploi lié à un traitement apparemment généreux, l'évaluation surestimera l'effet du traitement. De manière parallèle, l'expérimentation sur un petit groupe limite les effets de réseau et d'interaction sociale (*peer & network effects*). Par exemple, une personne ayant une information imparfaite des emplois potentiellement disponibles peut intensifier sa recherche d'emploi lorsqu'un proche retrouve un emploi (parce qu'elle réévaluera sa probabilité de trouver un emploi). Dans ce cas, un effet positif de court-terme sera moins marqué que l'effet de long-terme.

1.5. Solution aux effets d'équilibre général et dilemme des expérimentations

Revenons sur deux conditions de la validité interne d'une expérimentation : (1) les groupes de traitement et de contrôle doivent avoir des caractéristiques similaires et (2) le groupe de contrôle ne doit pas être affecté par le traitement. On peut montrer que dans une expérimentation sociale en environnement naturel, ces deux conditions sont difficiles à atteindre simultanément. Lorsqu'il existe des effets d'éviction, de contamination, ou d'équilibre général, la condition (2) n'est pas remplie. Sur le marché du travail, ces effets peuvent être élevés, notamment si on prend l'emploi comme proxy pour l'offre de travail (à cause des effets de file d'attente ou d'éviction) mais même si on mesurait la recherche d'emploi, le problème des effets de pairs subsisterait. Une personne peut se mettre à rechercher un emploi lorsque ses voisins ou connaissances se mettent à chercher ou trouvent un emploi. Les comportements économiques ne sont pas indépendants de normes sociales générées entre autre par le comportement des autres acteurs.

Une solution aux problèmes d'équilibre général est d'expérimenter au niveau où l'équilibre se forme. Concernant les politiques d'emploi, le niveau pertinent est probablement celui de la zone d'emploi. La solution consiste à traiter toutes les personnes dans une zone d'emploi donnée. Il ne peut alors y avoir d'éviction des personnes traitées par les personnes non traitées, puisque toutes les personnes dans la zone d'emploi sont traitées. On parle alors de site saturé (*saturation site*). La difficulté consiste alors à trouver un groupe de contrôle. Dans le cadre des politiques affectant l'emploi, le groupe de contrôle serait nécessairement une autre zone d'emploi. Mais, comparé à la technique de tirage aléatoire des groupes de contrôle et de traitement, la condition de similarité des deux groupes est moins bien remplie : il peut exister des différences de caractéristiques non aléatoires entre le groupe de contrôle et le groupe de traitement. Si ces différences sont non-observables, il sera impossible de contrôler pour ces différences de manière économétrique. De même, un choc affectant de manière asymétrique la zone traitée biaisera les résultats de l'expérience.

En présence d'effets d'équilibre général ou d'éviction, ce qui est vraisemblablement le cas concernant les politiques de l'emploi, l'expérimentation est nécessairement imparfaite. Ceci est en fait un problème commun aux sciences et particulièrement aux sciences sociales : l'expérimentation est l'objet d'un dilemme entre validité interne et externe. La validité interne suppose un contrôle important de l'environnement mais un tel contrôle pose le problème du caractère généralisable de l'expérience dans un univers non contrôlé. Dans le contexte de l'emploi, l'expérimentation contrôlée ne peut donc être considérée *a priori* comme un test plus sévère ou une méthode plus rigoureuse que les méthodes non expérimentales ou quasi-expérimentales. Selon l'expression de Cartwright (2007), l'expérimentation contrôlée ne peut être considérée comme « l'étalon-or » des méthodes scientifiques.

Nous nous proposons d'évaluer le protocole d'expérimentation du RSA à la lumière de ces difficultés méthodologiques.

2. Évaluer l'expérimentation du revenu de solidarité active

2.1. Le RSA : pérenne et familialisé

Allègre (2009) présente le RSA tel qu'il a été généralisé par rapport aux instruments visant à favoriser la reprise d'emploi des bénéficiaires de minima sociaux en place avant sa généralisation en juin 2009. Le RSA consiste à remplacer les minima sociaux (RMI, API) et leur système d'intéressement par une prestation dégressive et pérenne. Le mécanisme est conçu de telle sorte que toute progression des revenus d'activité se traduit par une diminution de la prestation inférieure à cette progression, afin que le revenu disponible du ménage progresse toujours lorsque ses revenus d'activité augmentent. Le RSA généralisé a un taux de cumul de 62 % (ou un taux de retrait de 38 %) : le RSA se réduit de 38 centimes et le revenu disponible du ménage augmente de 62 centimes – pour chaque euro d'activité

supplémentaire. Les autres revenus continuent à être déduits intégralement. L'allocation RSA généralisée est calculée comme suit :

$$\text{Allocation RSA} = \max (\text{allocation garantie} - \text{ressources hors activité du foyer} \\ - \text{Taux de retrait} * \text{revenus d'activité du foyer} ; 0)$$

L'allocation garantie est égale à celle existant précédemment sous le RMI et l'API. Elle dépend de la configuration familiale du foyer. L'allocation RSA peut être décomposée en un « RSA socle » qui serait égal à l'ancienne allocation RMI et un « RSA chapeau », complément de revenu d'activité. Le dispositif est fortement « familialisé » : son montant dépend fortement de la configuration familiale. Par exemple, un couple ayant des revenus d'activité mensuels équivalents à 1,5 SMIC (1 550 euros) ne percevra pas de RSA s'il n'a pas d'enfants et percevra une prime mensuelle de 115 euros s'il a un enfant et de 100 euros s'il a deux enfants, ce dernier couple recevant en sus des prestations familiales à hauteur de 125 euros. Le montant maximum de RSA chapeau est d'environ 440 euros, pour les couples avec 1 enfant, 425 euros pour les couples avec 2 enfants, 350 euros pour les célibataires avec 1 enfant et les couples sans enfant, mais inférieur à 250 euros pour les célibataires sans enfant.

Le caractère familialisé du RSA crée des effets ambigus en termes de gains financiers à la reprise d'emploi². Pour une personne seule, le gain du passage entre l'inactivité et le mi-temps au SMIC horaire est fortement augmenté : il est de 285 euros après la réforme RSA contre 145 euros avant la réforme, soit une augmentation du gain de 140 euros. Par contre, le gain du passage de l'inactivité au temps plein est réduit, ce qui se traduit par un taux marginal effectif d'imposition plus élevé dans cette zone : le gain n'est plus que de 175 euros contre 315 auparavant, soit une baisse de 140 euros. La réforme RSA ne changeant pas la situation des célibataires travaillant au SMIC à temps plein, l'augmentation des gains financiers pour la reprise d'emploi à mi-temps est entièrement compensée par la diminution de la différence de revenu disponible entre une activité à mi-temps et une activité à temps-plein au SMIC horaire. Pour les couples avec deux enfants, c'est le passage de l'inactivité à la monoactivité qui est favorisé par la réforme. La reprise d'emploi à temps plein au SMIC horaire du premier conjoint se traduit par une augmentation pérenne du revenu disponible de 557 euros contre 340 auparavant (soit une augmentation de 217 euros). Par contre, le passage de la monoactivité à la biactivité à temps-plein est moins rémunérateur : le gain pérenne est de 459 euros avec la réforme RSA contre 676 euros auparavant (soit une diminution de 217 euros). La réforme est donc ambiguë en termes d'incitation au travail : si pour les allocataires de minima sociaux, elle accroît les gains à la reprise à temps partiel pour les célibataires et au passage de l'inactivité à la monoactivité pour les couples ; pour les travailleurs pauvres, les gains liés à l'augmentation du nombre

2. Voir Allègre (2009) pour une présentation plus complète des modifications des contraintes budgétaires créées par l'introduction du RSA.

d'heures travaillées pour les célibataires et au passage de la monoactivité à la biactivité pour les couples sont réduits.

2.2. Cadre légal et champ de l'expérimentation

Selon l'article 18 de la loi TEPA « le revenu de solidarité active est mis en œuvre simultanément dans les conditions définies aux articles 19 et 20 de la présente loi pour les bénéficiaires du revenu minimum d'insertion et de l'allocation de parent isolé ». Le revenu de solidarité active expérimental est par ailleurs défini par le même article : « le revenu de solidarité active a pour objectif d'assurer l'augmentation des ressources d'une personne bénéficiaire d'un minimum social qui prend ou reprend un travail, exerce ou accroît son activité afin d'atteindre un revenu garanti qui tient compte des revenus d'activité professionnelle et des charges de famille ». L'article suivant stipule que « cette mise en œuvre est effectuée par les départements volontaires pour une durée de trois ans ». La loi de 2007 encadre donc strictement les expérimentations. Ne sont concernés que les bénéficiaires du RMI et de l'API et non tous les travailleurs pauvres comme le dispositif généralisé le prévoit. Ce sont les départements volontaires qui mettent en œuvre l'expérimentation et choisissent les modalités de sa mise en œuvre (notamment en termes d'accompagnement). Par contre, ils ne peuvent jouer que sur ce taux de cumul et non sur le montant de l'allocation garantie aux inactifs puisque la loi précise que seuls sont concernés les bénéficiaires d'un minimum social reprenant un emploi. Le montant de l'allocation doit également dépendre des ressources et des charges appréciées au niveau de la famille ; il ne peut être individualisé. L'expérimentation est donc à la fois décentralisée et très encadrée par rapport à ce qui peut être testé. Au final, ce n'est pas 1 mais 34 expériences contrôlées qui ont été menées. La loi prévoyant l'expérimentation est finalement abrogée au 1^{er} juin 2009 à la suite de la loi du 1^{er} décembre 2008 généralisant le revenu de solidarité active.

Limiter le champ de l'expérience aux bénéficiaires du RMI et de l'API et non à tous les bénéficiaires potentiels du dispositif généralisé pose plusieurs problèmes. Premièrement, le protocole ne permet pas de répondre à la question de la désincitation éventuelle du travailleur supplémentaire chez les couples. Seul l'impact de plus fortes incitations sur l'offre de travail et l'emploi des allocataires de minima sociaux est mesuré par l'expérimentation. Même dans un pur modèle d'offre de travail, l'impact global sur l'emploi de la mise en place du RSA généralisé ne peut donc pas être déduit des résultats de l'expérimentation, quelle que soit la significativité des résultats. De plus, le dispositif mesure un équilibre partiel et ne permet pas de répondre aux problèmes d'évictions. En effet, les allocataires de minima sociaux pourraient évincer de l'emploi les travailleurs peu qualifiés non-éligibles au dispositif expérimental. Du point de vue des politiques publiques, l'impact global sur l'emploi et plus précisément sur l'emploi en équivalent temps plein (en tenant compte de l'emploi à temps partiel) est important car il détermine à la fois le coût net *ex post* (après changement de comportement et bouclage fiscal) de la réforme, et l'impact du dispositif sur la pauvreté et les inégalités.

2.3. Territoires expérimentaux et territoires de contrôle

L'expérimentation du RSA n'est pas une pure expérience avec tirage aléatoire mais s'apparente à une quasi-expérience (Goujard, 2009). La loi ne permet pas de tirer au sort des individus pour composer les groupes de traitement et de contrôle mais donne la possibilité aux départements retenus d'expérimenter le RSA sur tout ou partie de leur territoire. Les conseils généraux retenus pour l'expérimentation ont donc choisi un territoire d'expérimentation (zone test) et les évaluateurs ont ensuite sélectionné des territoires de contrôles (zone témoin) aux caractéristiques proches des territoires tests.

Il est préférable d'un point de vue statistique d'affecter les individus aléatoirement dans les groupes de contrôle et les groupes de traitement plutôt que d'utiliser des zones tests et témoins. En effet, le nombre de zones est nécessairement beaucoup plus limité que celui d'individus et la présence d'effet fixe pour chaque zone réduit considérablement la puissance statistique de l'expérimentation. Dans le cas du RSA, les évaluateurs disposent de 85 zones (45 tests et 40 témoins) en fin de période, ce qui limite la précision statistique des effets à estimer (Gurgand et Montialoux, 2009). L'utilisation de territoires expérimentaux pourrait toutefois constituer une solution aux problèmes d'éviction et d'équilibre général présentés précédemment. Dans le contexte de l'expérimentation du RSA, l'utilisation de zones tests et témoins paraît de ce point de vue être un compromis satisfaisant. Les biais dus aux effets de débordement et d'éviction devraient être limités par l'utilisation de territoires expérimentaux qui correspondent plus ou moins à des zones d'emploi. Par contre, le prix à payer est une moins grande puissance statistique : du fait du nombre limité de zones, l'effet réel de l'expérimentation doit être très large (un taux de retour à l'emploi 20 % plus élevé) pour avoir une chance d'avoir un estimateur statistiquement significatif (Gurgand et Montialoux, 2009). Cette méthode de contrôle permet également d'éviter les problèmes éthiques et administratifs liés au tirage aléatoire des groupes expérimentaux.

En pratique, les zones tests et témoins n'ont pas été choisies de façon aléatoire, ce qui constitue une limite méthodologique. Les évaluateurs ont utilisé des critères quantitatifs et qualitatifs pour appairer une zone témoin à chaque zone test afin de minimiser le biais de sélection (Goujard et Lhorty, 2008) : (1) la forme administrative de la zone témoin devait être la même que celle de la zone test ; (2) la zone témoin devait être proche en termes de taille (population totale et nombre d'allocataires du RMI) de la zone test ; (3) les dynamiques d'emploi devaient être proches (« chances de sortir du RM et/ou celles de sortir du chômage, durée de chômage, ancienneté du RMI ») ; (4) « des réseaux d'acteurs comparables ou des 'effets territoires' comparables dans les deux zones ». Dans une première étape, les évaluateurs ont choisi des « zones témoins candidates » dont les caractéristiques sociodémographiques étaient comparables à celles de la zone test du point de vue de la taille, de la densité de population ou de la situation administrative. Puis, dans une deuxième étape, ils ont choisi parmi les zones candidates, la zone ayant, dans les

années précédant le démarrage de l'expérimentation, le taux d'entrée en emploi des allocataires du RMI le plus proche de celui de la zone test.

L'analyse des critères menant au choix des zones tests par les départements est déterminante pour évaluer la validité de l'expérimentation. Les évaluateurs n'ont eu accès qu'à un nombre limité de données administratives pour sélectionner des zones témoins. Si le choix des zones tests s'est fait par rapport à des données non observables par les évaluateurs et que ce choix n'est pas indépendant de la probabilité d'accès à l'emploi des allocataires de minima sociaux, il existerait un biais qui ne peut être contrôlé par des méthodes économétriques. Il est important de déterminer si ce biais éventuel peut mettre à mal la validité de l'expérimentation. *A priori*, les conseils généraux ont probablement des informations qui ne sont pas disponibles aux évaluateurs : ils peuvent par exemple savoir si des entreprises vont s'installer ou au contraire quitter la zone d'emploi, ou s'il existe un certain nombre d'entreprises mobilisables dans le cadre d'une politique volontariste d'accès à l'emploi des allocataires des minima sociaux. Les évaluateurs n'ont accès qu'à des informations rétrospectives sur les flux d'entrées et de sorties du chômage et du RMI. La question de la motivation des conseils généraux dans le choix des zones tests est ainsi cruciale pour la validité de l'expérimentation. Si les conseils généraux ont voulu faire en sorte que l'expérimentation ait un effet réel sur l'emploi, ils ont eu la possibilité de choisir des zones tests ayant des caractéristiques favorables. Au contraire, s'ils ont voulu garantir la validité scientifique de l'expérimentation, ils ont pu choisir des zones tests représentatives du département ou de façon aléatoire. Enfin, s'ils ont voulu faire bénéficier des territoires défavorisés d'une politique redistributive, ils ont pu choisir une zone test ayant des caractéristiques défavorables par rapport aux possibilités de retour à l'emploi des allocataires de minima sociaux.

Nous avons peu d'informations permettant de déterminer la motivation des conseils généraux dans le choix des zones tests. Les dossiers motivant le choix des zones tests par les départements ne sont pas inclus dans le rapport final de l'expérimentation. Loncle et *alii* (2009) ont mené une enquête monographique dans 5 départements sur la mise en œuvre de l'expérimentation du RSA. Ils soulignent plusieurs motivations pour le choix des territoires : la Côte d'Or et le Nord auraient choisi leurs territoires expérimentaux pour leur représentativité ; l'Eure et le Val d'Oise pour leur « exemplarité positive » (territoires dynamiques sur le plan économique) ; la Loire-Atlantique pour son « exemplarité négative » (territoire connaissant de nombreuses difficultés sociales). Dans ce sens, les motivations sont diverses et se compenseraient. Mais les auteurs ajoutent également que dans quatre des cinq cas, sont évoqués « des problèmes d'inadéquation entre l'offre et la demande d'emploi dans les territoires d'expérimentation » et que tous mentionnent la qualité du travail en réseau dans les territoires. Les conseils généraux semblent donc avoir choisi des territoires où il existe une demande d'emploi inassouvie, ce qui pourrait constituer une menace sérieuse à la validité de l'expérimentation. Le rapport final note également que « les départements ont mentionné dans les dossiers de candidatures au RSA, de façon presque

systématique, qu'ils se sont appuyés sur « un réseau d'acteurs de qualité ». Goujard et Lhorty (2008) concluent qu'« il est donc probable que les zones témoins aient des caractéristiques spécifiques qui facilitent le retour à l'emploi des allocataires du RSA ». L'ampleur du biais de sélection dépend de la motivation des conseils généraux, des informations à sa disposition, et des conséquences réelles des caractéristiques favorables sur le taux de retour à l'emploi. L'analyse de la motivation du choix des zones tests par chacun des conseils généraux serait nécessaire pour juger de la validité de l'expérimentation. La plupart des comptes-rendus des délibérations des conseils généraux concernant la mise en place de l'expérimentation du RSA ne donnent pas de justification du choix des territoires expérimentaux. Une analyse exhaustive des motivations des conseils généraux demanderait des recherches qui dépassent l'objet de cette étude. Les justifications données dans les comptes-rendus peuvent néanmoins donner un aperçu des critères de décision des conseils généraux. Allègre (2009) compile les justifications avancées par les 5 conseils généraux d'après le compte-rendu de leur délibération. Les conseils généraux semblent bien intégrer un principe de représentativité de la zone test (« un territoire à l'image du département » ; « ces territoires ont l'intérêt de présenter des caractéristiques socio-économiques diversifiées »). La nécessité d'un nombre suffisant d'allocataires du RMI est également mise en avant. Les conseils généraux utilisent ainsi des critères « scientifiques » (qui concernent la validité de l'expérimentation ou le caractère généralisable de ses leçons). En même temps, les conseils généraux ont voulu agir de façon efficace de telle sorte que les sommes dépensées au nom de l'incitation financière puissent avoir un réel impact sur l'emploi : l'Allier choisit « une zone dynamique en termes d'emploi et des secteurs d'activité ayant des difficultés de recrutement », la Charente souligne « la présence d'un réseau dynamique d'acteurs au service des personnes en difficulté » dans la zone choisie de sorte que ne soit pas proposés une incitation et un accompagnement dans l'emploi « dans un contexte où il n'y aurait pas de poste à pourvoir, ou bien des postes qu'elles [les personnes sans emploi] ne pourraient raisonnablement pas viser ». Les Côte d'Armor prennent en compte « une offre d'emploi existante et pour autant un nombre persistant de personnes en grande précarité ». D'autres conseils généraux ont choisi, probablement pour des raisons d'équité, d'expérimenter sur l'intégralité de leur territoire : c'est le cas de la Creuse, de la Haute Corse et de la Mayenne. Le département des Bouches-du-Rhône qui avait initialement envisagé une mise en œuvre sur l'intégralité du territoire a finalement choisi d'expérimenter sur l'ensemble de la ville de Marseille qui représente 70 % des bénéficiaires du RMI du département. Ces derniers critères peuvent être considérés comme étant politiques (au sens premier : relatif au gouvernement) dans la mesure où l'objectif n'est pas la validité scientifique de l'expérimentation mais la plus grande efficacité ou équité d'une politique publique mise en œuvre au niveau départemental. De même que Goujard et L'Horty, nous concluons qu'il est probable que les zones témoins aient des caractéristiques favorables du point de vue de l'accès à l'emploi, non seulement parce que les conseils généraux se sont appuyés sur des réseaux d'acteurs sociaux de qualité, mais surtout parce que, au nom d'un

motif d'efficacité de la dépense publique, ils ont choisi des zones où les incitations financières avaient le plus de chance d'être efficaces (les zones où il existe à la fois des postes et des personnes pouvant les occuper). Pour contrôler ce biais, les auteurs proposent de sélectionner les territoires témoins de façon « raisonnée » selon les principes énoncés ci-dessus. Il semble toutefois que le biais de sélection soit difficilement contrôlable par les évaluateurs.

La sélection de la ville de Marseille comme zone test dans les Bouches-du-Rhône illustre une deuxième difficulté liée au choix de la zone expérimentale par les conseils généraux, et donc en partie sur des considérations politiques. Il est en effet très difficile de trouver une zone de contrôle pour la ville de Marseille puisqu'il n'y a pas d'équivalent de la ville ni dans le département, ni dans la région. La zone témoin finalement choisie par les évaluateurs est le reste du département. Mais cette zone ne semble pas constituer un bon témoin : le nombre d'allocataires du RMI est deux fois plus faible que dans la zone test, la proportion d'adultes bénéficiaires du RMI y est 2,5 fois plus faible (4,94 % contre 12,61 %). Ceci peut poser un problème puisque le nombre d'allocataires du RMI sur la commune de Marseille correspond à 41 % du nombre d'allocataires du RMI dans l'ensemble des zones tests (voir Allègre 2009). Il semble bien exister une tension entre l'objectif scientifique des évaluateurs (l'évaluation sans biais d'un impact) et les objectifs d'efficacité et/ou d'équité pris en compte par les conseils généraux. Cette tension existait déjà, sous d'autres formes, dans les expérimentations d'incitation financière à la reprise d'emploi conduites en Amérique du Nord (Allègre, 2008).

L'ampleur du biais de sélection est difficile à évaluer puisqu'il dépend par définition de caractéristiques inobservables. Les évaluateurs peuvent contrôler le biais à l'aide de caractéristiques rétrospectives observables (notamment les flux d'entrées et de sorties des chômeurs et des bénéficiaires du RMI). Ils auraient pu exclure des départements de l'évaluation des résultats au motif d'un biais éventuel dans la sélection de la zone test ou de la difficulté liée à la sélection d'une zone témoin lorsque la zone test a des caractéristiques très spécifiques. Les évaluateurs ont plutôt choisi de faire jouer la loi des grands nombres : si certaines zones témoins ne sont pas satisfaisantes du point de vue des évaluateurs, au niveau agrégé, l'ensemble des zones témoins est jugé être un contrefactuel satisfaisant³. En agrégeant les résultats, les diverses motivations des conseils généraux se compenseraient. En pratique, les évolutions des zones tests et témoins entre 2001 et 2006 apparaissent semblables après agrégation et contrôle des caractéristiques observables (Goujard, 2009).

3. « Au niveau national, l'ensemble des zones témoins s'avère être un contrefactuel satisfaisant de l'ensemble des zones tests, même si, au niveau départemental ou pour une zone test donnée, les évolutions des zones témoins peuvent différer des territoires tests. » (Goujard, 2009).

2.4. Le traitement : RSA expérimental contre intéressement

L'effet mesuré dans une expérimentation est l'impact de la différence entre ce qui est donné au groupe de traitement (ici, l'accès potentiel au revenu de solidarité active expérimental) et ce qui est donné au groupe de contrôle (ici, l'accès à l'intéressement et à la prime de retour à l'emploi).

La partie a) du tableau 1 résume les gains types à la reprise d'emploi au SMIC pendant la première année pour des célibataires sans enfant du groupe de contrôle ayant droit à l'intéressement et à la prime de retour à l'emploi (PRE) et du groupe de traitement ayant droit au RSA avec un taux de cumul à 70 %. Le RSA étant pérenne, il est plus avantageux à long terme ; le tableau inclut également le nombre de mois nécessaires pour que la reprise d'emploi avec le RSA soit plus avantageuse qu'avec le système antérieur. Lors des trois premiers mois suivant la reprise d'emploi, les individus des deux groupes peuvent garder l'intégralité du RMI à condition d'être éligible à l'intéressement⁴. À l'issue de ces trois mois, les gains à la reprise diffèrent. Pour les reprises d'emploi de plus de 78 heures mensuelles (mi-temps), les individus du groupe non expérimental ont droit à une prime de retour à l'emploi égale à 1 000 euros lorsqu'ils reprennent une activité salariée qu'ils exercent pendant quatre mois consécutifs à condition de ne pas avoir déjà reçu cette prime dans les 18 mois précédent. Ils ont également droit à une prime forfaitaire d'intéressement (150 euros pour un célibataire) pendant 9 mois. Les individus du groupe de contrôle reçoivent eux une allocation qui dépend fortement du nombre d'heures travaillées : elle est plus importante pour les reprises d'emploi à mi-temps (235 euros) que pour les reprises à plein-temps (82 euros). Pour une reprise d'emploi à plein-temps, le RSA est plus favorable que le système d'intéressement à partir de 22 mois d'activité. Pour une reprise à mi-temps ce délai est de 13 mois. En ce qui concerne les reprises d'emploi de moins de 78 heures, le mécanisme expérimenté est plus favorable dès le troisième mois d'activité : il est non seulement pérenne, mais l'allocation mensuelle reçue est plus élevée car le taux de dégressivité de l'allocation est de 30 % contre 50 % pour le système d'intéressement. À court terme, la différence est néanmoins faible (moins de 40 euros mensuels pour un quart-temps). Pour les personnes seules, plus la reprise d'emploi est à temps long, moins le RSA expérimental est favorable par rapport à l'intéressement.

La partie b) du tableau 1 décrit la même situation pour une personne en couple avec deux enfants dont le conjoint est inactif et permet de juger l'effet de la familialisation des deux instruments. La prime forfaitaire d'intéressement est de 225 euros (au lieu de 150) ; du fait de sa familialisation, le RSA expérimental est plus favorable : les montants sont supérieurs dès la première année d'activité et quelle que soit la durée de reprise d'emploi. Ce sont pour les reprises à temps partiel long (autour d'un mi-temps), que le RSA expérimental est le plus favorable. La différence est de 144 euros mensuels pour un mi-temps (227 si la personne n'est pas

4. Les individus ayant déjà bénéficié d'un intéressement ont droit à un nouveau cycle complet d'intéressement de 12 mois après une interruption d'activité d'au moins 6 mois.

éligible à la PRE). Elle n'est que de 44 euros pour une reprise à quart-temps et de 29 euros pour une reprise à plein temps.

Tableau 1 : Prestations mensuelles moyennes la première année suivant la reprise d'emploi au SMIC pour les individus du groupe de contrôle et du groupe de traitement et nombre de mois d'activité nécessaires pour que le RSA expérimental soit favorable

a) Cas d'une personne seule

	Gain salarial	Intéressement + PRE	RSA (70%)	Mois
Reprise à quart-temps (9 h/semaine)	267	$(3*394+9*262)/12$ =295	$(3*394+9*313)/12$ =333	3,0
Reprise à mi-temps (18h/semaine)	533	$(3*394+1000+9*150)/12$ =294	$(3*394+9*235)/12$ =275	12,8
Reprise à plein-temps (35h/semaine)	1037	$(3*394+1000+9*150)/12$ =294	$(3*394+9*82)/12$ =160	22,1

b) Cas d'une personne vivant en couple avec deux enfants

	Gain salarial	Intéressement + PRE	RSA (70%)	Mois
Reprise à quart-temps (9 h/semaine)	267	$(3*394+9*552)/12$ =512	$(3*394+9*605)/12$ =552	3,0
Reprise à mi-temps (18h/semaine)	533	$(3*394+1000+9*225)/12$ =350	$(3*394+9*527)/12$ =494	6,7
Reprise à plein-temps (35h/semaine)	1037	$(3*394+1000+9*225)/12$ =350	$(3*394+9*374)/12$ =379	9,5

Source : Calculs de l'auteur.

2.5. Le caractère transitoire de l'expérimentation

La différence la plus importante entre le RSA et l'intéressement tel qu'il était pratiqué jusqu'en juin 2009 est l'aspect pérenne du RSA. Il est donc important de se demander si les travailleurs répondent à des incitations financières pérennes plutôt que transitoires. Ceci est particulièrement difficile à évaluer, notamment dans le cadre d'une expérimentation, qui par construction est à durée limitée. Si une durée longue est préférable du point de vue de l'évaluation, elle pose le problème de l'inégalité de traitement. En France, l'expérimentation n'est possible que pour une durée de 5 ans. Cette durée est relativement longue : elle excède probablement ce que les travailleurs anticipent en termes de stabilité des politiques publiques. Par exemple, les mesures d'intéressement ont fait l'objet de réformes

importantes en 1999 et 2006 avant d'être supprimées en 2009 ; la PPE n'a pas gardé les mêmes paramètres plus de deux années de suite depuis sa création en 2001. La difficulté liée à l'incertitude sur la durée d'une politique n'est donc pas propre à l'expérimentation. Elle peut toutefois peser plus lourdement lors d'une expérimentation.

La durée initiale des expérimentations prévues par la loi de finances 2007 et la loi TEPA était de 3 ans à compter de la publication du décret d'application autorisant la participation des collectivités territoriales volontaires. Ce décret a été publié en septembre 2007⁵ ; la fin légale de l'expérimentation était donc initialement prévue pour septembre 2010. Les expérimentations du RSA ont débuté entre novembre 2007 (Charente, Oise, Val d'Oise) et avril 2008 (Alpes Maritimes)⁶. La fin de l'expérimentation a été entérinée en décembre 2008 par la loi généralisant le RSA au 1^{er} juin 2009. L'expérimentation a finalement duré 17 mois en moyenne dont seulement 11 mois sans que les participants en connaissent officiellement le terme.

Les expériences étrangères démontrent l'importance de la durée du traitement lorsqu'on évalue les compléments de revenus. En effet, dans plusieurs expérimentations menées en Amérique du Nord, des effets initiaux importants sont devenus non-significatifs au cours de l'expérimentation (Allègre, 2008). Dans ces expérimentations, l'impact d'incitations financières à la reprise d'emploi plus importante a eu tendance à augmenter durant les premiers trimestres, peut-être à cause d'une meilleure information sur le dispositif et au délai lié à la recherche d'emploi (puisqu'on mesure l'emploi et non l'offre d'emploi). Ensuite, l'impact a eu tendance à se réduire, peut-être parce que les emplois occupés n'étaient finalement pas assez attractifs, ou parce que les travailleurs n'ont pas pu garder leur emploi (Allègre, 2008).

Si la généralisation du RSA a mis fin à l'expérimentation telle qu'elle était en place, elle n'empêchait pas de continuer à expérimenter l'impact des incitations financières avec les instruments déjà en place (zones tests et témoins, instruments de suivi...). Par exemple, un supplément au RSA aurait pu être mis en place dans les zones tests. On aurait également pu mettre en place un socle RSA plus élevé dans les zones tests associé à un taux de retrait plus élevé (de telle sorte qu'il n'y ait pas de perdants en zones tests, mais que les gains de retour à l'emploi soient plus faibles en zones tests qu'en zones témoins dans la seconde période de l'expérimentation) : la comparaison entre les deux périodes d'expérimentation aurait ainsi permis un bon contrôle du biais de sélection et des effets de mobilisation.

5. Le décret n°2007-1392 du 28 septembre 2007, complété par les décrets n°2007-1552 du 31 octobre 2007 et n° 2007-1879 du 26 décembre 2007, a autorisé 34 départements à participer à la mise en œuvre du revenu de solidarité active.

6. Le département de l'Eure a commencé son expérimentation en juin 2007 mais cette expérimentation ne relevait pas du même décret d'application (décret n°2007-691 du 4 mai 2007).

2.6. La question des effets de mobilisation

À l'exception de l'Eure, la communication sur le dispositif expérimental a été ciblée sur les bénéficiaires potentiels du RSA dans les territoires expérimentaux. Les conseils généraux ont redouté les critiques concernant les inégalités de traitement induits par les expérimentations (Comité d'évaluation des expérimentations, 2009). Ils ont donc privilégié l'information ciblée des allocataires potentiellement concernés par mailing ou réunion collective plutôt qu'une information générale à l'ensemble de la population du département.

Dans un rapport sur l'expérimentation, le directeur général adjoint du Conseil général du Val d'Oise souligne ces effets mobilisateurs : « L'expérimentation a agi, sur le site, mais également bien au-delà, comme un booster, comme un stimulant, de notre politique d'insertion. Sur le site d'expérimentation, nous avons été bien plus loin que les objectifs que nous nous étions donné : effort de contractualisation, accompagnement dans l'emploi, travail au côté des employeurs, travail sur le croisement entre formation et insertion, etc. » (ANSA, 2009). Selon le rapport final du comité d'évaluation, l'expérimentation a également eu pour effet une plus forte mobilisation des acteurs sociaux : « la mobilisation et la coopération de ces acteurs dans l'expérimentation sont plus ou moins inédites et plus ou moins renforcées ». Le rapport final note également que « quelques départements ont engagé des actions spécifiques de mobilisation des entreprises ciblées sur certains secteurs d'activité ou généralisées à l'ensemble des employeurs du bassin d'emploi (mailing, petits déjeuners avec les chefs d'entreprise, groupes de travail, démarchages dans les entreprises, forum emplois...). Le souhait de développer les liens avec les entreprises découle généralement de constats antérieurs à l'expérimentation RSA, sur l'inadéquation entre les offres d'emploi et les qualifications des personnes par exemple ».

Il semble donc que les acteurs ayant mis en œuvre l'expérimentation aient été peu sensibles à la question des effets d'une mobilisation asymétrique sur les territoires expérimentaux. Il est difficile d'évaluer l'impact potentiel sur les reprises d'emploi que ces effets peuvent avoir mais il est probable que cet impact soit décroissant dans le temps. Les effets de mobilisation pourraient ainsi également expliquer l'impact décroissant des dispositifs d'incitations financières dans les expérimentations nord-américaines. Il est dommage de ce point de vue que les expérimentations aient été interrompues prématurément.

Il serait intéressant de comparer la différence de taux d'entrée en emploi avant et après la généralisation du RSA en juin 2009 (différence-en-différence) : si le taux d'entrée en emploi reste supérieur en zones tests à celui des zones témoins après juin 2009, on pourra conclure que la différence n'est pas due au traitement mais soit à des effets de mobilisation, soit à des différences de caractéristiques entre zones tests et zones témoin. Dans une expérience avec tirage aléatoire, la différence-en-différence avant/après démarrage de l'expérimentation permet de contrôler pour la non-similarité des zones tests et témoins. Il ne permet néanmoins pas de contrôler des effets de mobilisation. De plus, dans le contexte de l'expérimentation du RSA, le

calcul de la différence-en différence n'est pas suffisant puisque les zones témoins ont été sélectionnées afin justement de minimiser la différence de taux d'entrée dans l'emploi avant démarrage de l'expérimentation. Le calcul de la différence après la fin de l'expérimentation permet de vérifier que la faible différence avant démarrage était bien due à une réelle similarité des zones. Il est donc important de continuer à observer les zones expérimentales après la fin de l'expérimentation. Le rapport final du comité d'évaluation ayant été publié en mai 2009, il n'inclut pas ce contrôle important. Ceci illustre une contradiction entre l'objectif institutionnel de l'expérimentation et son objectif scientifique. Selon la loi, la différence de traitement induite par l'expérimentation ne peut se justifier que par l'objectif de généralisation : les mesures expérimentées ont vocation à être étendues ou abandonnées à l'issue de l'expérimentation. La loi prévoit ainsi que les départements et le comité d'évaluation présentent un rapport d'évaluation au Gouvernement « avant l'expiration de la durée fixée pour l'expérimentation »⁷. Le Gouvernement doit également présenter un rapport au Parlement avant l'expiration de cette même durée. Or, l'objectif scientifique visant à vérifier l'hypothèse d'un impact causal des incitations financières sur le retour à l'emploi demande un temps plus long : il demande notamment d'observer les comportements après la phase expérimentale.

2.7. Le contexte économique

Le contexte économique n'a pas été favorable à l'évaluation d'incitations financières à la reprise d'emploi. L'expérimentation permet en théorie de contrôler les éléments de contexte : si les groupes de traitement et de contrôle sont similaires, ils sont affectés de la même manière par les évolutions du contexte économique. La méthode de différence-en-différence permet alors de supprimer l'effet propre du contexte économique. Mais le contexte économique peut également avoir un effet sur l'impact des incitations financières sur l'emploi. On peut facilement concevoir que l'importance des incitations financières soit moins élevée en période de chômage de masse que lorsque l'économie est proche du plein-emploi. En effet, pour que les incitations financières aient un effet sur l'emploi, il faut que les personnes augmentant leur offre de travail puissent réellement en trouver un. Ceci illustre le fait que les comportements économiques dépendent fortement de leur contexte, ce qui ajoute une limite supplémentaire au raisonnement expérimental : la causalité qu'une expérimentation sociale entend démontrer n'est en fait exacte qu'à un moment et à un endroit donné⁸.

Le graphique 1 présente le nombre de chômeurs au sens du BIT du 1^{er} trimestre 2007 au 1^{er} trimestre 2009 (prévision INSEE). Alors que le début des expérimentations (4^e trimestre 2007 et 1^{er} trimestre 2008), se déroule dans un contexte favorable du point de vue de l'emploi, la majeure partie des

7. Loi n°2006-1666 du 21 décembre 2006.

8. Ceci est vrai de toute recherche empirique en sciences sociales. Il convient de recontextualiser ces enseignements empiriques pour ne pas tomber dans « l'illusion expérimentale » (Passeron, 1991). Les causalités ne peuvent être déduites qu'après multiplication des observations.

expérimentations (du 1^{er} trimestre 2008 au 1^{er} trimestre 2009) se déroule dans un contexte d'augmentation rapide du chômage. Notons également que la crise économique a des effets très différenciés sur les secteurs puisqu'au début de la crise, le secteur industriel ainsi que le bâtiment ont été beaucoup plus nettement affectés par les pertes d'emploi que les services ou l'agriculture (Département analyse et prévision, 2009). Cet impact asymétrique de la crise augmente l'importance d'un bon contrôle. Il est important que la structure de l'emploi sectoriel soit similaire dans les zones tests et témoins. Dans le cas contraire, il est possible que les dynamiques de l'emploi dans les zones tests et témoins divergent pendant la crise alors qu'elles avaient connu des évolutions parallèles avant la crise.

Graphique 1 : Nombre de chômeurs au sens du BIT



Source : INSEE.

2.8. Résultats de l'expérimentation

Le rapport final du comité d'évaluation a été publié en mai 2009 (Comité d'évaluation des expérimentations, 2009). Outre l'analyse des effets sur l'emploi basée sur les données administratives, plusieurs études complémentaires ont été menées : la DRESS a réalisé une enquête quantitative auprès de bénéficiaires du RMI et de l'API résidant dans les zones tests et témoins (Fabre et Sautory, 2009) ; le CREDOC a mené une enquête qualitative à base d'entretiens individuels et collectifs auprès des personnes visées par le dispositif (Alberola *et alii*, 2009). Nous commentons ici les résultats en termes d'emploi.

Le résultat principalement commenté dans le rapport final du comité d'évaluation est l'impact des dispositifs expérimentés sur le taux d'entrée en emploi des allocataires de minima sociaux. En moyenne, le taux d'entrée en emploi dans les

zones expérimentales est supérieur à celui prévalent dans les zones témoins mais l'écart varie entre départements et selon les périodes. L'écart mensuel moyen entre zones expérimentales et zones témoins, égal à 0,28 point de pourcentage, n'est pas significatif au seuil de 5 %. Le taux mensuel moyen d'entrée dans l'emploi en zones témoins étant de 3,1 %, l'impact de 0,28 point correspond à un surcroît de 9 % avec un intervalle de confiance allant de - 6 % à 25 % (Comité d'évaluation des expérimentations, 2009). Ce taux de retour en emploi supérieur en zone expérimentale a un effet sur l'emploi : pour 100 bénéficiaires du RMI sans emploi en décembre 2007, on en compte 93,9 en zones expérimentales en décembre 2008 et 97,7 en zones témoins, soit un écart de 3,8 points (tableau 2).

Tableau 2 : Évolution du nombre de bénéficiaires du RMI sans emploi

Indice 100 en décembre 2007

	déc-07	mars-08	juin-08	sept-08	déc-08
Zone expérimentale RSA	100,0	100,5	96,3	93,0	93,9
Zone témoin	100,0	102,3	99,2	96,0	97,7

Source : Cnaf, Comité d'Évaluation des expérimentations (2009).

Ces résultats sont calculés par une méthode d'équipondération des départements. Les évaluateurs justifient ce choix par le poids très important de certains départements dans l'expérimentation (notamment les Bouches-du-Rhône et dans une moindre mesure l'Hérault) ainsi que par la diversité des dispositifs expérimentés. Notons que plus la taille de zone expérimentale est importante, plus il est difficile de trouver une zone de contrôle satisfaisante. De ce point de vue, le critère mis en avant par les conseils généraux concernant la taille suffisante de la zone expérimentale est ambigu du point de vue de l'évaluation : elle améliore la significativité statistique mais détériore la qualité du contrôle (Goujard, 2009). Étant donné que les résultats finaux sont équipondérés – les quelques 50 000 allocataires des Bouches-du-Rhône ont un poids équivalent aux 300 allocataires d'Ille et Vilaine – il aurait probablement été préférable que les zones expérimentales choisies soient de plus faible taille. L'équipondération des départements pose aussi le problème d'interprétation et de généralisation des résultats. La méthode donne une importance trop importante aux petits départements ruraux et en général aux zones où le nombre d'allocataires du RMI est faible. Deux zones de même population auront le même poids que la proportion d'allocataire du RMI y soit d'1 % ou de 10 % : l'équipondération revient à donner un poids dix fois plus élevé aux premiers. Si l'impact des dispositifs expérimentés sur le taux d'entrée en emploi dépend de la taille de la zone ou de la proportion d'allocataire du RMI, la moyenne équipondérée sera biaisée.

Des calculs pondérés par les effectifs expérimentaux ont également été effectués et confirment la sensibilité des résultats à l'inclusion ou non des Bouches-du-Rhône (Gurgand et Montialoux, 2009). En utilisant une pondération proportionnelle aux effectifs expérimentaux, l'écart de taux de retour en emploi entre zones tests et zone témoins n'est plus que de 0,03 points de pourcentage, soit un surcroît de taux de retour en emploi de seulement 1,1 %. Avec un écart-type égal à 10,1 %, ce surcroît est hautement non significatif). Si l'on exclut les Bouches-du-Rhône du calcul pondéré, l'écart de taux de retour en emploi est de 0,34 points, soit un surcroît toujours non-significatif mais qui s'élève alors à 14 % (avec un écart-type égal à 10,3 %). La sensibilité du résultat n'est pas seulement liée au poids des Bouches-du-Rhône mais également à l'hétérogénéité des résultats dans les différents départements (si le résultat des Bouches-du-Rhône était proche du résultat moyen des autres départements, l'effet de son inclusion ou non ne serait pas aussi important). La publication des résultats détaillés par département et par couple zone test/zone témoin permettrait de se faire une meilleure idée de cette hétérogénéité et de juger de la pertinence du choix d'agréger les résultats ainsi que de celui de la méthode de pondération.

Gurgand et Montialoux (2009) évaluent la comparabilité des zones tests et témoins en calculant une différence de taux d'entrée à l'emploi avant le démarrage du RSA. Les auteurs utilisent les données de déclarations trimestrielles de ressources datant de janvier 2008 à mars 2008 pour les départements démarrant l'expérimentation après ces dates. Ces données n'étaient donc pas disponibles lors de la sélection des zones témoins. Sur les 22 départements observés, l'écart est de 0,31 points de pourcentage avant démarrage du RSA. Cet écart est équivalent à celui observé après démarrage des expérimentations sur l'ensemble des départements (0,28 points). Néanmoins, l'écart de 0,31 est presque entièrement dû au seul département de Charente-Maritime : en excluant ce département, l'écart n'est plus que de 0,06 points et la différence-en-différence sur les 21 départements est égal à 0,24 points⁹. En excluant le département de Charente-Maritime et en raisonnant en différence-en-différence avant/après démarrage de l'expérimentation, le surcroît de taux d'entrée en emploi dans les zones tests est de 7,6 %.

Afin de compléter les résultats issus de données administratives, la DREES a mené une enquête quantitative entre mai et juin 2008 auprès des personnes résidant dans les zones tests et témoins qui étaient bénéficiaires du RMI ou de l'API au 31 octobre 2007 (Fabre et Sautory, 2009). Les auteurs concluent également à un impact limité en termes d'accès ou de maintien à l'emploi. Parmi les personnes sans emploi le mois précédent le démarrage des expérimentations, 14,3 % étaient en emploi dans les zones tests en mai 2008 contre 14,0 % dans les zones témoins, soit un écart non significatif de 0,3 points (tableau 3). L'écart concernant les personnes ayant commencé un emploi au cours des expérimentations est de 1,5 point (également non significatif au seuil de 90 %). Un plus grand nombre de personnes a

9. Avec un écart-type de 0,26 points, l'écart est non-significatif.

commencé un emploi au cours de l'expérimentation sans se maintenir en emploi jusqu'en mai 2008 en zone tests qu'en zone témoins (4,8 % contre 3,6 %, soit un écart d'1,2 points). En théorie l'effet du dispositif sur le maintien en emploi des personnes retrouvant un emploi en cours d'expérimentation est ambigu : d'une part le gain financier devrait inciter des personnes à accepter des emplois précaires mais d'autre part, pour les personnes qui auraient de toute façon accepté un emploi l'incitation financière devrait avoir un impact positif sur le maintien en emploi (comme le montre l'écart d'1,3 points entre zones tests et témoins dans la proportion des personnes en emploi avant expérimentations ayant conservé un emploi en mai 2008). Les résultats suggèrent donc que l'effet négatif dû aux caractéristiques des emplois acceptés dominerait l'effet positif.

Tableau 3 : Impact de l'expérimentation sur l'accès à l'emploi

En %

	Ensemble	Zones tests	Zones témoins	Écart* (en points)
Sans emploi avant expérimentations				
– a commencé un emploi	18,4	19,1	17,6	1,5
– est en emploi en mai 2008	14,1	14,3	14,0	0,3
En emploi avant expérimentations**				
– a conservé son emploi en mai 2008	71,2	71,7	70,4	1,3

* Non significatifs au seuil de 90 % .

** Dans les zones reprenant le stock de personnes en emploi.

Source : Enquête sur l'expérimentation du RSA, DREES ; Fabre et Sautory (2009).

L'enquête quantitative montre également que les emplois occupés par les individus en zones tests sont de plus faible durée et à moindre rémunération mensuelle qu'en zones témoins. Il y a significativement plus de personnes en emploi de moins de 9h en zones tests (10,6 %) qu'en zones témoins (5,8 %) et significativement moins de personnes percevant 1 000 euros ou plus de salaires mensuels (31,4 % contre 39,7 %). À part l'effet sur la durée de travail, il y a peu de différences entre les emplois recherchés ou occupés des deux zones (Fabre et Sautory, 2009) : les jugements sur la stabilité de l'emploi, les horaires, les salaires et l'adéquation du travail aux qualifications sont semblables entre zones tests et zones témoins. Néanmoins et de façon inattendue, les conditions de travail sont jugées plus satisfaisantes en zones tests qu'en zones témoins : 83 % des personnes ayant commencé leur emploi après le démarrage des expérimentations sont satisfaites de leurs conditions de travail en zones tests contre 73 % en zones témoins. On s'attend au contraire à ce que les emplois aient de moins bonnes conditions de travail dans les zones où le gain financier est plus important (puisque certaines personnes acceptent des emplois qu'ils refusaient ou ne cherchaient pas auparavant). Comment expliquer ce paradoxe ? Remarquons d'abord que le taux d'emploi n'est pas

beaucoup plus élevé en zones tests : on s'attendrait donc à un très faible effet négatif sur les conditions de travail. Ensuite, le jugement sur les conditions de travail peut-être affecté par la perspective du gain financier. Enfin, le retour à l'emploi en zones tests se fait en moyenne sur de plus faibles durées, ce qui peut rendre plus acceptables des conditions de travail similaires par ailleurs. Quoiqu'il en soit, on n'observe pas de conditions de travail dégradées en zones tests.

3. Expérimenter avant de généraliser ?

Les expérimentations ont finalement apporté relativement peu d'enseignements sur l'efficacité des mécanismes d'incitation à l'emploi. Les évaluations d'impact sur l'emploi des dispositifs testés sont non significatives. On pourrait conclure que les effets estimés sont faibles mais la puissance statistique de l'expérimentation était elle-même très faible et les dispositifs évalués étaient peu différents en termes de gain à la reprise d'emploi que le droit commun, du moins à court terme.

La différence essentielle entre le dispositif évalué (le RSA) et le droit commun (l'intéressement) est l'aspect pérenne du RSA. Mais l'expérimentation n'a finalement duré que 17 mois en moyenne dont seulement 11 mois en moyenne sans que les participants en connaissent le terme. Comme les gains financiers à la reprise ne sont réellement différents dans le dispositif évalué qu'après 12 mois, il apparaît que le protocole mis en place ne permet pas de répondre à la question de l'impact d'incitations pérennes à l'emploi. Gomel et Serverin (2009) concluent que « l'expérimentation ne porte pas sur « l'augmentation des ressources » (...) mais sur l'efficacité incitative de la promesse d'un accroissement de ressources sur la reprise d'un travail ». L'impact mesuré serait celui d'un discours sur les incitations plutôt que l'impact des incitations elles-mêmes.

De plus, les différences de taux d'entrée en emploi entre zones tests et zones témoins pourraient être dues à d'autres facteurs, notamment des différences de dynamiques d'emploi entre zones tests et zones témoins dues à une sélection non-aléatoire des zones expérimentales. Il est difficile d'évaluer l'ampleur de ces biais mais, étant donné la faiblesse de l'impact du traitement sur le taux d'entrée en emploi (le taux moyen mensuel d'entrée dans l'emploi est supérieur de 0,28 point en zones tests), il n'est pas improbable que celui-ci soit entièrement dû à des facteurs autres que l'incitation financière. Par conséquent, on ne peut réellement conclure dans un sens ou dans l'autre sur l'existence et l'ampleur d'un impact de gains monétaires pérennes sur la reprise d'emploi des allocataires de minima sociaux.

Enfin, le protocole mis en place ne permettait pas de répondre à de nombreuses autres questions soulevées par la mise en place d'un complément de revenu familialisé pour travailleurs pauvres. En effet, le dispositif expérimenté diffère du dispositif généralisé. Notamment, les travailleurs pauvres non bénéficiaires de minima sociaux n'étaient pas éligibles au dispositif expérimental alors que la principale spécificité du RSA est de bénéficier à tous les travailleurs pauvres qu'ils soient passés ou non par les minima sociaux.

L'échec relatif de l'expérimentation s'explique en partie par l'hétérogénéité des objectifs poursuivis par les acteurs qui l'ont mise en place. D'un point de vue scientifique, la méthode expérimentale permet de tester la validité d'une hypothèse. Ici, il s'agit de vérifier si un dispositif de type RSA, et plus généralement la mise en place d'incitations financières pérennes à l'emploi, ont un effet sur la reprise d'emploi des allocataires de minima sociaux. L'expérimentation est alors un succès si elle permet de répondre à cette question et s'il est possible de mesurer sans biais et le plus précisément possible l'ampleur de cet effet. Or, la mesure a été imprécise et les biais ne sont potentiellement pas négligeables étant donnée l'ampleur de l'effet mesuré. Les conseils généraux ont été peu sensibilisés aux objectifs scientifiques et notamment au fait que la mesure doit se faire sans biais. Au contraire, il semble qu'un certain nombre de conseils généraux ait utilisé le RSA expérimental comme un instrument supplémentaire dans les outils d'insertion et de lutte contre la pauvreté. L'objectif de lutte contre l'exclusion et la pauvreté poursuivi par certains conseils généraux est en partie contradictoire avec l'objectif scientifique de mesure sans biais de l'impact du dispositif expérimenté¹⁰.

La loi française n'autorise l'expérimentation sociale que dans l'objectif d'une généralisation. Mais, l'expérimentation contrôlée produit des résultats en équilibres partiels qui apportent le plus souvent peu d'enseignements sur les effets d'une généralisation éventuelle : quel effet sur l'emploi ? Sur le chômage ? Sur le nombre d'allocataires de minima sociaux ? Quel coût global de la mesure ? En 2005, Lise *et alii* tentent de répondre à ces questions pour le projet d'autosuffisance canadien qui s'est déroulé entre 1992 et 1995. Il apparaît clairement que temps scientifique et décisions publiques sont peu compatibles : alors que les décideurs ont besoin d'une réponse avant l'issue de l'expérimentation afin de généraliser ou non le protocole expérimental, le consensus scientifique met des années à s'établir. Pour qu'il puisse s'établir, un débat scientifique est nécessaire. Or, les expérimentations contrôlées, notamment les expérimentations de grande ampleur, sont peu propices à l'établissement d'un tel débat : l'expérimentation n'est pas reproductible, les données ne sont le plus souvent pas disponibles à toute la communauté scientifique. De plus, du point de vue de l'évaluation de la politique publique, l'expérimentation a tendance à réduire le champ du débat démocratique qui se focalise alors sur ce que tente de mesurer l'expérimentation (ici le niveau de l'emploi), et à négliger les autres effets (réduction de la pauvreté, précarisation éventuelle de l'emploi, etc.). L'expérimentation sociale peut aussi avoir l'inconvénient d'évincer les autres approches du débat public : à notre connaissance, aucune évaluation de l'impact du RSA sur l'emploi se basant sur un modèle structurel d'offre de travail n'a été publiée avant la généralisation du dispositif.

L'expérimentation du RSA est caractéristique de l'ambiguïté des expérimentations sociales de grande ampleur. Elles poursuivent deux objectifs :

10. C'est pourquoi l'expérimentation pose des questions éthiques et est souvent rejetée par les acteurs sur le terrain, notamment lorsque les groupes expérimentaux sont tirés au sort.

celui d'évaluer une politique publique et celui de tester la validité scientifique d'une théorie. Nous soutenons que, dans le domaine des incitations financières à l'emploi, les expérimentations sociales n'ont jusqu'ici pas permis de répondre à ces questions de manière satisfaisante. Ces deux objectifs (évaluer une politique publique, tester une théorie) sont en partie contradictoires et soulignent le dilemme des expérimentations entre validité interne et validité externe. La validité interne nécessaire au test de la théorie suppose un contrôle important de l'environnement mais un tel contrôle pose le problème du caractère généralisable de l'expérience dans un univers non contrôlé, nécessaire pour l'évaluation de la politique publique.

Références bibliographiques

- Alberola E., M Angotti et M. Brezault, 2009 : *Enquête qualitative auprès de bénéficiaires visés par le RSA*, CREDOC.
- Allègre G., 2008, « L'expérimentation sociale des incitations financières à l'emploi : questions méthodologiques et leçons des expériences nord-américaines », *Document de travail OFCE*, n° 2008-22.
- Allègre G., 2009, « L'expérimentation du revenu de solidarité active entre objectifs scientifiques et politiques », *Document de travail OFCE*, n° 2009-36.
- Blundell R. et I. Walker, 2001, « Working Families' Tax Credit: A Review of the Evidence, Issues and Prospects for Further Research », *Inland Revenue Research Report*, n°1.
- Blundell R., M. Costa Dias et C. Meghir, 2003, « The Impact of Wage Subsidies: A General Equilibrium Approach », *Unpublished manuscript*, Institute for Fiscal Studies.
- Bourguignon F. et A. Chiappori, 1998, « Fiscalité et redistribution », *Revue française d'économie*, vol 12.
- Burtless G. et D. Greenberg, 1982, « Inferences Concerning Labor Supply Behavior Based on Limited-Duration Experiments », *The American Economic Review*, Vol. 72, n° 3.
- Burtless G. et L. Orr, 1986, « Are Classical Experiments Needed for Manpower Policy? », *The Journal of Human Resources*, Vol. 21, n° 4.
- Cahuc P., 2002, « À quoi sert la prime pour l'emploi ? », *Revue française d'économie*, Volume 16, n° 3.
- Card D. et P. Robins, 1996, « Do financial incentives encourage welfare recipients to work ? Evidence from a Randomized Evaluation of the Self-Sufficiency Project », *NBER Working Paper Series*, n° 5701.
- Cartwright N., 2007, « Are RCTs the Gold Standard? », *Technical Report, Centre for Philosophy of Natural and Social Science*, The London School of Economics and Political Science.
- Clement P., 2002, « Rapport sur le projet de loi constitutionnelle relatif à l'organisation décentralisée de la République », Commission des lois constitutionnelles, *Rapport de l'Assemblée Nationale*, n° 376.

- Comité d'évaluation des expérimentations, 2008, *Rapport d'étape sur l'évaluation des expérimentations RSA*, septembre.
- Comité d'évaluation des expérimentations, 2009, *Rapport final sur l'évaluation des expérimentations RSA*, mai.
- CSERC, 1997, « Minima Sociaux, entre protection et insertion », *La documentation française*, Paris.
- Daubresse M.-P., 2008, *Rapport sur le projet de loi généralisant le revenu de solidarité active et réformant les politiques d'insertion*, Commission des affaires culturelles, familiales et sociales, Rapport de l'Assemblée Nationale, n° 1113.
- Département analyse et prevision, 2009, « France : la mer se retire. Perspectives 2009-2010 pour l'économie Française, OFCE, *Revue de l'OFCE*, n° 109.
- Fabre V. et O. Sautory, 2009, « Enquête sur les expérimentations du RSA : Premiers résultats », *Document de travail DREES*, Série Études et Recherches, n° 87.
- Fisher R.A., 1966, *The design of Experiments*, Oliver and Boyd.
- Godino R., 1999, « Pour une réforme du RMI », *Notes de la fondation Saint-Simon*, n° 104.
- Gomel B. et E. Serverin, 2009, « Expérimenter pour décider ? Le RSA en débat », Document de travail IRERP, n° 119.
- Goujard A. et Y. Lhorty, 2008, « Définition des zones témoins pour l'expérimentation du revenu de Solidarité active (rSa) », *Rapport de recherche CEE*, n° 43
- Goujard A., 2009, *Comparaison des zones tests et témoins avant expérimentation*, Groupe de recherche SOLSTICE, Centre d'Études de l'Emploi dans *Rapport final sur l'évaluation des expérimentations RSA*, Comité d'évaluation des expérimentations.
- Gurgand M. et C. Montialoux, 2009, « Impact du RSA sur les taux d'entrée en emploi des bénéficiaires du RMI », *Note préparée pour le rapport du Comité national d'évaluation du RSA*.
- Heckman J.J., Smith J.A., 1995, « Assessing the Case of Social Experiments », *Journal of Economic Perspectives*, vol. 9, n° 2.
- Hirsch M., 2005, *Au Possible nous sommes tenus*, ministère des Solidarités, de la Santé et de la Famille. Commission familles, vulnérabilité, pauvreté. La documentation Française, Paris.
- Lise J., S. Seitz et J. Smith, 2005, « Equilibrium Policy Experiments and the Evaluation of Social Programs », *Working Papers* 1076, Queen's University, Department of Economics.
- Loncle P., V. Muniglia et T. Rivard, 2009, *La mise en œuvre de l'expérimentation du RSA*, École des Hautes Études en Santé Publique (EHESP), département ShScS-LERFAS dans *Rapport final sur l'évaluation des expérimentations RSA*, Comité d'évaluation des expérimentations.
- Metcalf C., 1974, « Predicting the Effects of Permanent Programs from a Limited Duration Experiment », *The Journal of Human Resources*, Vol. 9, n° 4.
- Mirrlees J., 1971, « An Exploration in the Theory of Optimum Income Taxation », *Review of Economic Studies*, vol. 38(114).
- Passeron J.-C., 1991, *Le raisonnement sociologique. L'espace non-popperien du raisonnement naturel*, Nathan, Paris.

■ Guillaume Allègre

- Piketty T., 1998, « L'impact des incitations financières au travail sur les comportements individuels : une estimation pour le cas français », *Économie et prévision*, no 132-133.
- Robbins P., 1983, « The Labor Supply Response of Twenty-Year Families in the Denver Income Maintenance Experiment », *The Review of Economics and Statistics*, Vol. 66, n° 3.
- Rubin D., 1974, « Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies », *Journal of Educational Psychology*, n° 66.