



Document de travail

« L'expérimentation sociale des incitations financières à l'emploi : questions méthodologiques et leçons des expériences nord-américaines »

**N° 2008-22
Juillet 2008**

Guillaume Allègre
(OFCE-Centre de recherche en économie de Sciences Po)

Observatoire Français des Conjonctures Économiques

69, Quai d'Orsay 75340 Paris Cedex 07

Tel : 01 44 18 54 00 Fax : 01 45 56 06 15

E-mail: ofce@ofce.sciences-po.fr Web: <http://www.ofce.sciences-po.fr>

L'expérimentation sociale des incitations financières à l'emploi : questions méthodologiques et leçons des expériences nord-américaines

Guillaume Allègre
OFCE - Centre de recherche en économie de Sciences Po

2008-22
Juillet 2008

Introduction

L'expérimentation du revenu de solidarité active (RSA) prévue dans la loi du 21 août 2007 en faveur du travail, de l'emploi et du pouvoir d'achat (« loi Tèpe ») est la première expérimentation sociale de grande ampleur en France. La loi prévoit la mise en œuvre du RSA à titre expérimental dans les départements volontaires pour une durée de trois ans. Dans le cadre de cette expérimentation, les allocataires de minima sociaux se voient garantir une augmentation de leur ressources en cas de reprise d'activité et ce dès les premières heures travaillées. Les minima sociaux passent ainsi d'une logique différentielle (où le montant de la prestation diminue d'un euro lorsque les revenus d'activité augmentent d'un euro) à une logique dégressive (où le montant de la prestation diminue moins vite que la hausse des revenus d'activité). D'un point de vue scientifique, l'objet de l'expérimentation est d'évaluer l'effet sur l'offre de travail des allocataires de minima sociaux de plus fortes incitations financières à l'emploi.

L'expérimentation sociale est plus développée dans les pays anglo-saxons ainsi que certains pays d'Europe du Nord qu'en France. C'est dans les années 1960 et 1970 aux Etats-Unis que se sont déroulées les premières expérimentations sociales dans le domaine économique. Le *New Jersey Experiment* mis en place en 1968 est souvent présenté comme étant la première expérience contrôlée sociale de grande ampleur. Elle fait figure aujourd'hui d'expérimentation sur l'expérimentation sociale. L'objectif était à la fois d'évaluer l'impact sur l'offre de travail de la mise en place d'un revenu minimum garanti couplé à un impôt négatif et de démontrer la pertinence de l'expérimentation sociale dans l'évaluation des politiques publiques. Suite à cette expérimentation, trois autres expériences similaires ont été conduites aux Etats-Unis. Plus récemment, le Projet d'autosuffisance a été mené à partir de 1992 au Canada afin d'évaluer l'effet d'une augmentation importante des gains monétaires à l'emploi pour les bénéficiaires de l'aide sociale. De même que pour les expériences américaines d'impôt négatif, le Projet d'autosuffisance fait l'objet d'une littérature abondante. Le *Minnesota Family Investment Program* (MFIP) mis en place entre 1994 et 1998 est l'expérimentation la plus proche de celle du RSA puisqu'il s'agit de transformer un minimum social d'une allocation différentielle en une allocation dégressive. D'autres expérimentations menées aux Etats-Unis dans les années 1990 ont tenté d'évaluer l'effet de suppléments de revenu d'activité sur l'emploi des prestataires d'aide sociale.

Le Projet d'autosuffisance canadien est souvent cité comme preuve de réussite des programmes d'incitations financières à l'emploi. Les impacts produits sur l'emploi à plein-temps y sont très importants. Nous montrons que le programme évalué est très particulier et que les résultats ne peuvent être généralisés. Surtout, les résultats en termes d'emploi de l'ensemble des expérimentations sur les incitations financières menées aux États-Unis et au Canada depuis les années 1970 sont beaucoup plus ambigus qu'une lecture rapide des résultats du Projet d'autosuffisance ne pourrait laisser penser. Globalement, ces résultats sont conformes à ceux obtenus par des méthodes non-expérimentales.

Une première partie est consacrée aux éléments méthodologiques des expérimentations sociales. Si les expérimentations peuvent résoudre le problème du biais de sélection auquel sont confrontées les méthodes non expérimentales, il existe en pratique de nombreux obstacles qui sont susceptibles de menacer la validité scientifique de telles expériences. La portée de ces menaces dans le contexte des expérimentations des incitations financières à l'emploi est discutée. Dans une seconde partie, nous présentons les protocoles et les résultats des expériences conduites en Amérique du Nord visant à modifier les incitations financières pour les bénéficiaires de l'aide sociale. Nous concluons sur les leçons que nous pouvons en tirer dans le cadre de l'expérimentation du revenu de solidarité active.

Atouts et limites de l'expérimentation sociale dans le cadre de l'évaluation des incitations financières

Solution au problème de biais de sélection et lisibilité des résultats

Les évaluations empiriques des élasticités d'offre de travail font l'objet de nombreuses controverses. L'évaluation de l'impact des incitations financières se heurte en effet à la mesure des gains monétaires à l'emploi : nous ne pouvons pas 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 - où le salaire dépend de caractéristiques observées pertinentes (âge, années d'études, années d'expérience...) - 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 sont très différentes : l'estimation de l'équation de salaire sur les personnes en emploi fait l'objet d'un biais de sélection. Le salaire auquel un individu peut prétendre dépend en fait d'un ensemble de caractéristiques que l'on n'observe pas dans les enquêtes (en partie parce qu'elles ne sont pas quantifiables). Si les personnes sans emploi ont des caractéristiques inobservées défavorables, leur salaire potentiel sera inférieur au salaire estimé sur les personnes en emploi. Or, il paraît naturel de penser que les personnes sans emploi (notamment les chômeurs) ont des caractéristiques défavorables. Il existe des méthodes économétriques plus ou moins sophistiquées dont l'objectif est de réduire ce biais de sélection mais il est impossible de l'éliminer complètement sur données non expérimentales (Heckman et Smith, 1995).

L'expérimentation est, en théorie, une solution simple au problème de biais de sélection. Comme le soulignent Heckman et Smith, 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. Puisqu'il est impossible d'observer un individu à un moment donné dans deux états différents, le principe de l'expérimentation sociale 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. Ce principe est le même que celui des expérimentations médicales dont on utilise 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 ». Le traitement est alors la cause de la différence de comportement et de résultat des deux groupes sous deux conditions : (1) les deux groupes doivent être effectivement identiques et (2) le traitement doit être la seule circonstance qui diffère entre les deux groupes. Si ces deux conditions ne sont pas remplies, la validité interne de l'expérience n'est pas respectée et l'estimation de l'effet causal entre traitement et résultat 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 déjà randomisées, (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.

L'expérimentation permet ainsi de lever le doute sur la direction de la causalité entre traitement et résultats : il est toujours difficile de répondre à la question du sens la causalité en analysant des données non expérimentales (Burtless, 1995). Si la validité interne de l'expérience est assurée, on peut conclure que le traitement est la cause des différences de résultats et que les réactions différentes ne peuvent pas être la cause des différences de traitement.

Outre la validité scientifique, un précieux avantage de l'évaluation expérimentale est la facilité d'interprétation de ses résultats. Le protocole est plus facilement compréhensible que ne peuvent l'être les études sur données non expérimentales. La simplicité de l'analyse des résultats permet 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. L'expérimentation permet également d'évaluer des programmes qui n'ont pas été mis en place auparavant soit totalement nouveaux, soit d'une ampleur sans précédent. Dans ce cas, l'extrapolation de données non expérimentales peut engendrer des estimations biaisées.

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. A 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 le plus souvent bénéfique de façon non ambiguë. Dans le cadre des incitations financières à l'emploi, cela peut se traduire par un transfert financier vers les personnes qui trouvent un emploi ou augmentent leur nombre d'heures travaillées, si on veut tester l'effet d'une augmentation des incitations financières ou alors un transfert financier vers les personnes sans emploi si on veut évaluer l'effet d'une baisse des incitations financières¹. Dans ce cas, on refuse un avantage monétaire à des personnes du groupe de contrôle alors qu'elle est reçue 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

1. On peut également vouloir tester l'augmentation conjointe du revenu sans emploi et du revenu d'activité qui laisserait l'incitation financière à reprendre un emploi inchangée et qui permettrait ainsi de tester l'existence éventuelle d'un « effet revenu » chez les personnes peu qualifiées.

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.

En France, la loi exige lors des expérimentations biomédicales, qui peuvent avoir des conséquences beaucoup plus graves que les expérimentations sociales, qu'un document d'information complet sur les tenants et les aboutissants de la recherche soit remis aux intéressés et que leur consentement par écrit soit recueilli. Les expérimentations sociales sont également encadrées par la loi : elles doivent avoir un objet et une durée limités et être menée en vue d'une généralisation. Ce sont les collectivités locales qui se portent candidates à l'expérimentation : elles peuvent déroger aux dispositions législatives mais n'en ont pas l'obligation. Selon la loi constitutionnelle de 2003, l'expérimentation ne peut pas remettre en cause « les conditions essentielles d'exercice d'une liberté publique ou d'un droit constitutionnellement garanti »². La loi prévoit également l'obligation d'évaluation des expérimentations : avant l'expiration de l'expérience, un rapport exposant les effets des mesures expérimentées doit être présenté au Parlement³. Au contraire des expériences biomédicales, la loi ne prévoit pas d'autres protections des personnes. Elle est en outre silencieuse sur la possibilité d'expériences contrôlées avec tirage au sort.

Menaces à 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 lorsque (1) 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 et lorsque (2) le traitement à évaluer n'est pas le seul paramètre qui diffère entre les deux groupes.

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 auxquels 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

2. LOI constitutionnelle n° 2003-276 du 28 mars 2003 relative à l'organisation décentralisée de la République.

3. Cf. LOI organique n° 2003-704 du 1er août 2003 relative à l'expérimentation par les collectivités territoriales.

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 doit également 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 également changer de comportement parce qu'ils se savent l'objet d'une expérimentation. On parle d'« 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 quelques 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 peut biaiser 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 contrôlées biomédicales sont, lorsque c'est possible, effectuées en double aveugle : ni les patients, ni l'équipe médicale ne savent qui utilisent les produits actifs et qui utilisent les placebos. L'expérience en double aveugle est plus difficile en expérimentation sociale. Elle est impossible dans le cadre de l'évaluation des incitations financières : dans ce cas, le groupe de traitement doit nécessairement être informé 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 auxquels 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.

Menaces à 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 invariant 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 au nombre de personnes auxquelles il a été donné. Dans une expérimentation, le groupe de traitement jouit de l'avantage d'être seul dans la population à recevoir le traitement. L'effet du traitement peut être réduit, voire même s'annuler lorsqu'il est généralisé à la population entière s'il existe un phénomène dit de « file d'attente ». Il y a éviction lorsque le bénéficiaire du traitement se fait au détriment des personnes non traitées : elles sont dépassées par les personnes traitées dans une file d'attente dont la vitesse est fixe. En période de chômage de masse, des instruments ciblés sur certaines catégories de publics peuvent avoir des effets importants sur l'emploi de ce groupe et avoir des effets nuls au niveau national. 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 au dépend des personnes n'ayant pas accès à ce type de programmes. Il peut également jouer dans le cadre des incitations financières. Pour résoudre le problème lié aux files d'attente, une solution est de traiter toutes les personnes susceptibles d'être dépassées dans la file d'attente. Par exemple, lorsqu'il s'agit de mesurer un impact sur l'emploi, traiter toutes les personnes dans une zone d'emploi donnée permet de résoudre le problème de substitution. Mais dans ce cas, le groupe de contrôle ne peut être choisi de façon aléatoire dans l'échantillon. Le groupe de contrôle doit être choisi dans une ou plusieurs autres zones d'emplois. Cela affecte la validité interne de l'expérience puisqu'il existe des différences de caractéristiques non aléatoires entre le groupe de contrôle et le groupe de traitement. Un choc affectant de manière asymétrique la zone traitée biaisera les résultats de l'expérience. En présence d'effets de file d'attente, il existe donc un dilemme difficile à résoudre entre validité interne et validité externe de l'expérience.

De manière plus générale, les expérimentations sont limitées par le fait qu'elles ne mesurent qu'un équilibre partiel. Par construction, elles ne se situent pas au niveau de l'équilibre général. Par exemple, une expérimentation peut mesurer l'effet d'un programme sur les comportements d'offre de travail au niveau individuel. Mais dans un programme national, une modification de l'offre de travail aura des répercussions sur le salaire et donc la demande de travail : l'effet sur l'emploi dépendra également de l'élasticité de la demande. En théorie, dans le cas extrême où la demande serait totalement inélastique, une augmentation de l'offre de travail n'aurait pas de conséquence sur l'emploi et aurait pour seul effet soit une baisse du salaire soit une hausse du chômage.

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 (si par exemple, on augmente le revenu minimum garanti), 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'expérimentation de l'introduction ou de l'augmentation 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 l'expérience du New Jersey en utilisant des informations sur les comportements de consommation des foyers. Une autre solution pour évaluer ce biais consiste est à expérimenter plusieurs durées du traitement (Burless 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.

Une réelle simplicité d'interprétation ?

Le nombre important de biais auxquels les expériences contrôlées sont potentiellement confrontées oblige à reposer la question de la simplicité d'interprétation des résultats. Selon Heckman et Smith (1995), cet argument confond simplicité apparente et simplicité réelle : en présence de biais (d'échantillonnage, d'attrition, d'éviction...), les estimations des impacts expérimentaux peuvent être aussi difficile à interpréter honnêtement que les résultats d'études non expérimentales.

Si en théorie, l'expérimentation est une solution au problème du biais de sélection, il existe en pratique de nombreuses limites à ce type d'évaluation. Outre les limites méthodologiques et éthiques, les expérimentations peuvent être coûteuses en temps et en argent. Il existe aussi un risque que les délais scientifiques et politiques ne coïncident pas : il peut facilement se passer plusieurs années entre le lancement d'une expérimentation et la publication de résultats scientifiques. De façon plus générale, une expérimentation sociale de grande ampleur est nécessairement le résultat d'un processus politique. Or, les objectifs politiques et scientifiques

peuvent diverger, notamment lorsqu'il s'agit de mettre en place un protocole expérimental. L'objectif scientifique est alors, à coût administratif donné, de minimiser les biais qui peuvent affecter l'estimation de l'effet du traitement. La prise en compte des aspects politiques peut ajouter d'autres contraintes : les pouvoirs publics en charge de l'expérimentation voudront minimiser les éventuels coûts politiques liés aux questions éthiques. L'expérimentation peut être perçue comme un moyen d'augmenter l'acceptabilité politique d'une réforme. Cet aspect 'démonstration politique' peut affecter la rigueur scientifique de l'évaluation. On ne peut alors juger de l'opportunité scientifique d'une expérimentation sociale qu'en analysant l'impact des interactions entre contraintes scientifiques et politiques sur la définition et la mise en place du protocole expérimental.

Les expérimentations nord-américaines sur l'incitation financière au travail.

Quelles leçons peut-on tirer des expériences nord-américaines en matière d'incitation financière au travail ? Nous présentons ici plusieurs expérimentations significatives : les expériences d'impôt négatif qui ont été mises en place entre 1968 et 1971 aux États-Unis, le Projet d'autosuffisance (PAS) canadien qui se déroula entre 1992 et 1999 ainsi que les expérimentations mises en place dans les années 90 par le MDRC.

Dans cette partie, nous resituons le contexte politique et économique de ces expériences. Nous en présentons les objectifs scientifiques et les protocoles mis en place. Nous discutons des résultats et des critiques qui ont été émises. Nous nous limitons ici à commenter les résultats en termes d'emploi et de revenus. Les évaluations ont souvent mesuré d'autres effets : divorces et mises en couple, consommation, performance scolaire des enfants (voir Rossi et Lyall, 1976 ; Munnell, 1986 ; Gennetian et Miller, 2000)

Les premières expériences d'impôt négatif aux États-Unis

Contexte politique et économique

En 1964, le Président Lyndon Johnson déclare la « Guerre contre la pauvreté » (*War on Poverty*) dans son discours sur l'État de l'union. Afin de poursuivre cet objectif, il crée l'*Office of Economic Opportunity* (OEO) dont la mission est de développer et de gérer de nouveaux programmes sociaux. Au début des années 60, les programmes sociaux fédéraux les plus importants sont destinés aux personnes éloignées du marché du travail soit parce qu'elles sont trop âgées (*Old Age Assistance*), handicapées (*Aid to the Permanently and Totally Disabled*), ou qu'elles sont parents isolés (*Aid to Families with Dependent Children*, AFDC). Ce dernier programme avait été instauré par le *Social Security Act* en 1935 et bénéficie essentiellement aux mères isolées. A la création de l'AFDC, on n'attendait pas de celles-ci qu'elles travaillent. Au début des années 60, l'AFDC est décriée : comme les autres aides sociales, c'est une aide différentielle qui assure un revenu minimum garanti aux personnes dont les revenus sont inférieurs à ce niveau. Cette prestation crée par conséquent des taux de prélèvement effectifs égaux à 100 % : jusqu'au niveau du revenu minimum garanti, les revenus du travail viennent se soustraire aux transferts perçus et le gain monétaire à l'emploi est nul. Les critiques reprochent à cette aide d'être désincitative au travail. Ils lui reprochent également de favoriser les ruptures conjugales et les naissances hors mariage ainsi que son caractère stigmatisant. Les critiques sont alimentées par le fait que le nombre d'allocataires et le coût des aides sociales sont en constante augmentation.

L'idée de l'instauration d'un impôt négatif trouve son origine dans un livre de Friedman (1962) : il y propose de remplacer les multiples aides sociales existantes souvent gérées par les États (dont les aides alimentaires, aides au logement...) par un système fédéral où la protection sociale serait intégrée au système fiscal. Dans ce système, toute famille aurait droit à un revenu minimum (rm) fonction de la taille du foyer auquel serait soustrait une fraction des revenus du travail selon un taux de retrait (t). Tant que cette fraction est inférieure au revenu garanti, l'impôt est négatif : une prime est versée par l'État à la famille. A un certain niveau de revenus du travail (égal à rm/t), l'impôt négatif s'annule. Dans le système intégré de Friedman, au-delà de ce niveau, il devient positif sans créer d'effets de seuil. Outre les incitations au travail, Friedman souligne cinq autres avantages d'un tel système (d'après

Moffit, 2003) : (1) il se base uniquement sur le revenu, la taille et la composition du foyer et non sur les caractéristiques et le statut des personnes qui le compose (sexe, statut marital...) et serait donc plus équitable ; (2) les transferts monétaires sont préférables aux aides en nature (telles les aides alimentaires) du point de vue du bénéficiaire; (3) ce système serait plus simple, plus transparent et moins coûteux d'un point de vue administratif que les multiples aides existantes; (4) la redistribution serait plus facilement ciblée sur les personnes les plus pauvres; (5) contrairement à une hausse du salaire minimum, un tel système est efficace car il n'empêche pas la réalisation de l'équilibre de marché. Ce type d'instrument est le plus souvent appelé impôt négatif (*negative income tax*) mais dans un système socio-fiscal dépourvu de revenu minimum (*guaranteed minimum income*), la caractéristique la plus importante de la réforme proposée est en fait l'introduction d'un revenu minimum pour tous les foyers sans ressources.

L'OEO s'approprie rapidement l'idée d'un revenu minimum universel couplé d'un impôt négatif (Rossi et Lyall, 1976). La version d'impôt négatif de l'OEO a pour vocation d'universaliser le bénéfice de l'AFDC mais ne se substituerait pas à l'ensemble des dispositifs sociaux. Toutes les familles pourraient ainsi bénéficier du système social et non pas seulement les personnes éloignées du marché du travail telles les mères isolées. Si cette version de l'impôt négatif s'inspire des travaux de Friedman, l'esprit de la réforme en est très éloigné. Se pose alors la question de l'effet potentiellement négatif d'une telle mesure sur l'offre de travail des hommes d'âge actif. Ces derniers étaient à l'époque peu ou pas concernés par les programmes sociaux. Un revenu minimum universel réduirait leur incitation financière au travail : dans ce contexte, un nombre significatif d'hommes arrêteraient-ils de travailler ? Si c'était le cas, le coût du programme serait beaucoup plus important. Les défenseurs du revenu minimum universel savaient également qu'il serait difficile de convaincre l'opinion publique de la pertinence d'un tel instrument à moins de montrer que l'effet sur l'emploi des hommes était faible. Il y avait peu de données non expérimentales sur l'offre de travail des bas revenus et les études existantes donnaient des résultats très contrastés. Le principe d'une expérimentation sociale contrôlée est adopté par l'OEO. La motivation principale de l'expérimentation est politique : il s'agissait de démontrer que la mise en place d'un revenu minimum universel aurait peu de conséquences négatives sur l'emploi des hommes d'âge actif, qu'ils n'allaient pas « s'arrêter de travailler en masse » (Aaron, 1975). L'expérimentation sociale permettait également de rendre ce type de réforme plus visible aux yeux du public et des décideurs politiques⁴

L'expérimentation du New Jersey s'est concentrée sur l'étude du comportement des hommes d'âges actifs. Deux raisons principales ont mené à cette décision (Skidmore, 1975). En premier lieu, les hommes potentiellement actifs (non étudiant, non retraité, non handicapé) étaient le seul groupe à ne jamais avoir été éligible à l'assistance publique, il était donc difficile de prédire leur réaction potentielle à l'introduction d'un minimum social. En deuxième lieu, puisque peu de femmes travaillaient, les évaluateurs trouvaient probable qu'une baisse de l'offre de travail des femmes suite à l'introduction n'entraînerait pas de coûts supplémentaires significatifs ; un des objectifs principaux de l'expérimentation étant d'évaluer le coût supplémentaire dû à la baisse potentielle des taux d'emploi.

Suite à l'expérimentation du New Jersey, de nouvelles expérimentations sont lancées afin de palier à certaines insuffisances de cette première expérience. Alors que l'expérience dans le

4. Lors des élections de 1966, le Congrès passe sous le contrôle des républicains. L'OEO décide alors de faire référence aux incitations à l'emploi plutôt qu'à l'impôt négatif dans le nom du programme qui devient officiellement *New Jersey Graduated Work Incentive Experiment* (Levine, 1975).

New Jersey ne concernait que des populations urbaines, le *Rural Experiment* entend étudier le comportement des individus dans deux Etats ruraux du Midwest (Iowa) et du Sud (Caroline du Nord). Le *Gary Experiment* étudie la réaction de familles noires vivant dans un « ghetto ». Enfin, l'expérience à Seattle et Denver fut la dernière et la plus importante des expériences menées aux Etats-Unis à cette époque. L'échantillon de 5000 familles était de taille plus grande que celui des trois précédentes expériences et l'éligibilité n'était pas limitée par race, type de famille ou niveau de revenu précédant l'expérience.

Protocoles mis en place

Le tableau 1 décrit les protocoles mis en place dans les quatre expérimentations. Les sites expérimentaux ont été choisis parce qu'ils représentaient différents territoires de pauvreté : anciennement industriel (New Jersey), rural, ghetto urbain (Gary), et urbain (Seattle-Denver). Les premières expériences ont utilisé des critères d'éligibilité en termes de revenu relativement faibles (150% du seuil de pauvreté dans le New Jersey et le Rural experiment). Ce seuil fut critiqué par les problèmes d'endogénéité qu'il créait : avec un seuil de revenu peu élevé, les personnes éligibles sont celles qui sont par construction les moins attachées au marché du travail. Par conséquent, les expérimentateurs à Gary et à Seattle-Denver ont choisi des seuils plus élevés (respectivement 240% et 325% du seuil de pauvreté). Les expériences plus tardives ont également étudié l'impact d'autres mesures pouvant faciliter la reprise d'emploi (accompagnement, garde d'enfants). Les résultats en termes d'impact sur l'offre de travail des premières expérimentations étant faibles et non significatifs, les expérimentateurs ont choisi des traitements à Seattle-Denver plus généreux que dans les expériences antérieures. L'expérimentation de Seattle-Denver permettait également d'évaluer le biais de durée limitée en expérimentant plusieurs durées de traitement (3 et 5 ans).

Chacune des quatre expérimentations a testé plusieurs formes d'impôt négatif qui se différenciaient par le niveau du revenu minimum (rm) garanti aux foyers sans revenus d'activité et le taux de retrait (t) auquel l'aide est retirée lorsque les revenus d'activité s'accroissent. Dans le modèle néo-classique d'offre de travail, l'effet de l'introduction d'un impôt négatif sur l'offre de travail est négatif. Le revenu minimum augmente le revenu hors travail des agents : leur probabilité de cesser complètement de travailler est plus élevée. De plus la dégressivité de la prestation accroît le taux de prélèvement effectif : dans ce système, il est moins avantageux de travailler une heure de plus (puisque les revenus supplémentaires sont taxés à un taux t). Les agents sont alors incités à réduire le nombre d'heures travaillées. Ces deux effets (revenu, substitution) jouent donc dans le sens d'une réduction de l'offre de travail. L'intensité de ces effets dépend du niveau retenu pour le revenu minimum (rm) et le taux de retrait (t). L'objectif des expérimentations était également de mesurer les différences d'impact selon les niveaux rm et t retenus.

Tableau 1 : Protocoles mis en place

	New Jersey	Rural Experiment	Gary Experiment	Seattle-Denver
Sites	Trenton, Patterson-Pasaic et Jersey City, NJ ; Scraton, PA	Contés ruraux, Iowa et Caroline du Nord	Gary, Indiana	Seattle, Washington Denver, Colorado
Critères de sélection de l'échantillon	(1) Foyers comprenant un homme âgé de 18 à 58 ans et potentiellement actif ; (2) revenu inférieur à 150 % du seuil de pauvreté ; (3) le foyer devait comprendre au moins deux membres.	(1) Foyers comprenant au moins une personne à charge (2) revenu moyen inférieur à 150 % du seuil de pauvreté ;	(1) Foyers noirs dont le chef est âgé de 18 à 58 ans ; (2) Foyers comprenant au moins une personne à charge ; (3) revenu inférieur à 240 % du seuil de pauvreté	(1) Foyers comprenant au moins une personne à charge (2) revenu moyen inférieur à 325 % du seuil de pauvreté
Nombre de familles dans l'échantillon	1357 foyers ; 725 expérimentaux et 632 contrôles	809 foyers ;	1800 foyers ;	4801 foyers ;
Traitements	Huit combinaisons [rm, t]: [50, 30] ; [50, 50] ; [50, 70] ; [75, 30] ; [75, 50] ; [75, 70] ; [100, 50] ; [100, 70] et [125, 50].	Huit combinaisons [rm, t]:	* Quatre combinaisons [rm,t]: [40, 75] ; [60, 75] ; [40,100] ; [60,100]. * Accompagnement *Subvention garde d'enfants	*Onze combinaisons [rm,t]: [95, 50] ; [95, 70] ; [120,50] ; [120,70] ; [140,50] ; [140,70] *Accompagnement *Subvention formation
Durée de l'expérimentation	3 ans	3 ans	3 ans	3 ans, 5 ans (20 ans)
Début de l'expérimentation	1968	1970	1971	1971

Résultats : Impacts sur l'offre de travail

Les résultats des quatre expérimentations sont résumés dans le tableau 2 (d'après Robins, 1985). Chaque expérimentation a produit plusieurs estimations de l'impact moyen sur l'offre de travail. Mais les différents traitements ayant produits des impacts très similaires, l'auteur choisit de ne présenter qu'un résultat par genre et par expérience en se basant sur les traitements les plus similaires. Les impacts sur le taux d'emploi et le nombre d'heures travaillées dans l'année sont tous du signe attendu : l'introduction d'un impôt négatif avec revenu minimum réduit l'offre de travail. Néanmoins, seuls les résultats de Seattle-Denver sont significatifs : l'échantillon y était plus important et les traitements plus généreux. L'impact sur l'emploi des femmes en couple (- 8 %) est deux fois plus important que celui sur les hommes en couple (- 4 %).

Tableau 2 : Impact moyen sur l'offre de travail

Expérimentation	Hommes mariés		Femmes mariées	
	taux d'emploi	nombre annuel d'heures travaillées	taux d'emploi	nombre annuel d'heures travaillées
New Jersey	-0.01	-34.2	-0.4	-55.7
Rural	-0.003	-56	-0.16	-178
Gary	-0.01 (0.03)	-35.4 (65.1)	-0.03 (.04)	-57.6 (62.7)
Seattle-Denver	-0.04*** (0.01)	-128.6*** (30.1)	-0.08*** (0.2)	-145.1*** (31.4)

*** Résultat significatif au seuil de 1%

Source : Robins (1985)

L'expérimentation de plusieurs durées de traitement à Seattle et à Denver devait apporter une réponse sur l'importance du biais de durée limitée : 70 % des foyers ont été enrôlés dans un programme de 3 ans, les autres dans des programmes de 5 et 20 ans. Burtless et Greenberg (1982) comparent les réactions des foyers enrôlés dans les programmes de 3 et 5 ans. Les auteurs trouvent un impact de l'introduction de l'impôt négatif plus important dans le programme de 5 ans que dans le programme de 3 ans, cet effet étant plus marqué chez les hommes que les femmes. Ces résultats suggèrent que l'effet permanent de l'introduction d'un tel instrument est potentiellement plus important que lors d'une expérience à durée limitée. Les auteurs estiment un modèle structurel d'offre de travail séparément pour les individus enrôlés dans les programmes de 3 et 5 ans. Conformément aux prédictions théoriques (Metcalf, 1974), ils trouvent un effet revenu plus faible et un effet de substitution plus élevé dans le programme de 3 ans que dans le programme de 5 ans. Les auteurs ajoutent que si dans le cadre de l'expérience à Seattle et Denver, l'impact global est plus important dans le programme de 5 ans, d'autres protocoles - en particulier des traitements moins généreux - peuvent produire des résultats inverses. Robbins (1983) analyse les résultats du programme de 20 ans dans lequel 169 foyers ont été enrôlé et les compare à ceux des programmes de 3 et 5 ans. L'auteur ne trouve pas d'impact significatif sur l'offre de travail des personnes en couple du programme de 20 ans. Par contre, pour les personnes seules, l'impact est similaire à celui trouvé dans les programmes de 3 et 5 ans. Mais l'auteur montre que cet impact est croissant durant la période d'observation, ce qui suggère qu'il existe un effet d'apprentissage relativement élevé : il est nécessaire de suivre les foyers sur de longues périodes afin d'estimer les réponses à long-terme d'un changement de politique fiscale ou sociale.

Leçons des expérimentations

Les protocoles mis en place présentent de nombreuses limites soulignées par Rossi et Lyall (1976), Aaron (1975) et Widerquist (2005). Dans les premières expérimentations, seule la réaction des foyers les moins qualifiés a été évaluée : l'échantillon n'a pas été choisi au hasard mais selon un niveau de revenu. Ce niveau a été fixé très bas et de nombreux foyers qui auraient été éligibles au programme dans le cadre d'une généralisation, n'étaient pas éligibles au traitement : le traitement le plus généreux prévoyait un revenu minimum équivalent à 125 % du seuil de pauvreté et un taux de retrait de 50 %. Les foyers ayant des revenus inférieurs à 250 % (125 %/50 %) du seuil de pauvreté auraient été éligibles au programme. En se concentrant sur les personnes ayant des revenus inférieurs à 150 % du seuil de pauvreté, l'expérimentation a exclu de nombreux foyers de son analyse. Par conséquent les résultats peuvent difficilement être généralisés aux autres catégories de population. Plus grave, ce critère d'éligibilité crée un biais de sélection manifeste : le critère de revenu au niveau du foyer exclut pratiquement tous couples biactifs. Seules 16% des femmes comprises dans l'échantillon sont en emploi, contre plus de 30 % des femmes dans la population. Par conséquent, les femmes et les hommes inclus dans l'analyse ne sont pas représentatifs : du fait d'appariements sélectifs, les personnes dont le conjoint ne travaille pas n'ont pas les mêmes caractéristiques que celles dont le conjoint travaille. Ce problème a été en partie résolu dans l'expérimentation de Seattle-Denver où les expérimentateurs ont choisi un seuil d'éligibilité à l'expérimentation beaucoup plus élevé (325 % du seuil de pauvreté).

Le fait de choisir de conduire l'expérimentation sur des sites tests pose le problème de sa généralisation à l'ensemble du pays. Le choix de sites ayant des caractéristiques différentes répond en partie à ce souci. Une expérimentation nationale aurait permis une meilleure représentativité dans le contexte d'une évaluation de la généralisation de l'expérience au niveau fédéral.

Mahoney et Mahoney (1975) soulignent une autre limite de l'expérimentation. L'expérience a été conçue et conduite par des partisans de l'impôt négatif : les traitements proposés étaient donc naturellement différentes formes d'impôt négatif. Les évaluateurs partaient de l'hypothèse que des taux de prélèvements effectifs proches de 100 % avaient un fort effet désincitatif. Au vu des résultats qui montrent un faible effet des traitements, il est dommage que cette hypothèse n'ait pas été testée.

Les scientifiques ayant conduit cette expérience ont conclu qu'il y avait peu d'effets négatifs à l'introduction d'un NIT. Selon Hum et Simpson (1993), résumées en moins de 25 mots, les leçons des expérimentations seraient : « peu d'effets négatifs ont été trouvés jusqu'ici. Ceux trouvés, comme la réponse en termes d'offre de travail, sont plus petits qu'attendus sans expérimentation »⁵. Pourtant la perception médiatique et politique de cette expérimentation a été très différente. La plupart des rapports et des articles de presse ont insisté sur le fait qu'il y avait des effets désincitatifs et non pas qu'ils étaient faibles et souvent non-significatifs (Widerquist, 2005). Finalement l'effet voulu fut contraire à celui souhaité par les évaluateurs alors même que les résultats montraient une désincitation plus faible qu'attendue : au lieu de convaincre les décideurs publics du bien-fondé d'un impôt négatif, l'expérimentation a donné des arguments à ses opposants. Néanmoins, selon Coyle et Wildavsky (1986), l'interprétation des résultats des expérimentations n'est pas la cause principale de l'échec législatif d'un programme de revenu minimum / impôt négatif. Selon les auteurs, l'échec est plutôt la

5. *Few adverse effects have been found to date. Those adverse effects found, such as work response, are smaller than would have been expected without experimentation*

conséquence d'une absence de compromis entre députés égalitariens et libéraux face aux conservateurs : les égalitariens demandaient un niveau de revenu minimum beaucoup trop élevé pour que la réforme soit acceptée par les libéraux.

Selon Greenberg et Robbins (1986), l'expérimentation est par essence une force conservatrice : en concentrant l'attention sur les problèmes méthodologiques par nature complexes, l'expérimentation crée des controverses plutôt que du consensus. Les expérimentations évaluant l'impact sur l'emploi de l'introduction d'un impôt négatif ont le grand inconvénient de concentrer l'attention sur un effet secondaire non souhaité (l'impact négatif sur l'emploi) et non plus sur l'objectif principal (la réduction de la pauvreté). Reichauer (cité par Widerquist, 2005) pose la question de ce qu'il serait advenu de la Sécurité Sociale⁶ si elle avait été précédée par une expérimentation. Une expérience aurait certainement montré une baisse de l'épargne et un départ plus précoce à la retraite⁷. D'un point de vue politique, l'expérimentation comporte donc des risques pour les décideurs voulant mettre en œuvre une réforme, à moins de connaître par avance les effets du traitement. Mais dans ce cas, l'intérêt scientifique de l'expérimentation est limité.

Le Projet d'autosuffisance (PAS)

Contexte politique et économique

Le projet d'autosuffisance est lancé en 1992, 20 ans après les expériences d'impôt négatif aux États-Unis. Alors que ces expérimentations testaient les désincitations éventuelles de l'introduction d'un minimum social, le projet d'autosuffisance renverse le problème : un programme consistant à donner une prime aux personnes retrouvant un emploi pourrait-il inciter les allocataires d'aide sociale à travailler ?

Dans les quinze années précédant l'expérimentation, le coût du programme canadien d'aide au revenu (AR) triple (Mijanovich et Long, 1995). Le nombre d'allocataires croît également de façon importante. Parallèlement, les décideurs publics se préoccupent de plus en plus des effets potentiellement désincitatifs des programmes d'aide sociale. De même que l'AFDC, l'AR est une prestation différentielle : à moins de gagner plus que le montant de la prestation, les revenus du travail sont déduits des prestations reçues et les gains à l'emploi sont nuls. Mijanovich et Long donne l'exemple d'une personne célibataire avec deux enfants habitant dans le New Brunswick : son revenu annuel hors travail est de \$11,850 grâce aux versements d'AR. En travaillant 40 heures par semaine à un taux horaire de \$7⁸, son revenu serait de \$14,250 soit \$2,400 de plus qu'en ne travaillant pas. Chaque heure travaillée lui rapporte \$1,20 sans tenir compte des coûts éventuels de garde d'enfant et de transport. Travailler 10, 20 ou 30 heures ne changerait pas sa situation financière. De fait, seuls 10% des allocataires d'AR déclarent des revenus du travail (alors qu'ils n'ont pas d'intérêt monétaire à travailler). Le débat sur les politiques publiques s'est donc concentré sur le fait que le travail ne payait pas, notamment pour les parents isolés qui bénéficient d'une aide plus importante et dont les coûts liés à la reprise d'emploi sont plus élevés. Au début des années 90, un consensus politique s'établit pour réformer le système afin d'augmenter les incitations financières au travail. L'idée du PAS est d'expérimenter un programme d'incitations financières pour les allocataires de l'AR retrouvant un emploi à plein temps avant une généralisation éventuelle.

6. Instaurant une retraite publique par répartition.

7. Par contre, l'expérimentation peut avoir des effets bénéfiques sur la qualité du débat public.

8. \$1 de plus que le salaire minimum.

Protocole mis en place

Les évaluateurs ont tiré au sort un pré-échantillon parmi les foyers éligibles résidant dans certaines parties de Colombie Britannique et du New Brunswick. Ils devaient remplir les critères d'éligibilité suivants : (1) être parent isolé, (2) avoir 19 ans ou plus et (3) avoir reçu des versements d'AR pendant au moins 12 des 13 derniers mois. Après un entretien présentant le principe de l'expérimentation, 90 % des foyers ont donné leur accord pour y participer. Ces 6,028 parents isolés constituant l'échantillon ont alors été affectés de manière aléatoire aux groupes traitement et contrôle. Chacun avait une probabilité égale de rejoindre l'un ou l'autre groupe.

Afin de pouvoir bénéficier d'un supplément de revenu (SR), les membres du groupe de traitement devaient travailler au moins 120 heures sur une période de 4 semaines consécutives dans un ou plusieurs emplois. L'idée du SR est d'inciter la reprise d'un emploi à temps plein, condition nécessaire pour préparer l'autosuffisance.

Les personnes bénéficiant du SR voient leur revenu augmenter de façon substantielle. Le supplément de revenu dépend des revenus individuels et ne varie pas selon les charges de la famille : il est égal à la moitié de la différence entre les revenus d'activité du participant et un revenu de référence (\$30,000 au New Brunswick). Le supplément de revenu est donc maximum pour une personne travaillant 30 heures au salaire minimum. Il est ensuite réduit de 50 cents par dollar supplémentaire de revenus du travail. Reprenons l'exemple d'une mère célibataire reprenant le travail 40 heures par semaine pour un revenu du travail annuel égal à \$14,250. Elle recevrait \$7,875⁹ de supplément de revenu ce qui porterait son revenu à \$22,125. En travaillant 30 heures, son revenu du travail serait de \$10,900 et le supplément de revenu égal à \$9,500. Son revenu total serait donc de \$20,400. Une personne faisant partie du groupe de contrôle ayant les mêmes revenus du travail continuerait à bénéficier de l'AR et n'aurait aucun gain à la reprise d'emploi : son revenu serait toujours égal à \$11,850. Dans ce cas, faire partie du groupe de traitement permettait une augmentation du revenu de plus de 70 %.

Une personne pouvait bénéficier du supplément de revenu pendant 3 ans à partir de la reprise d'emploi. Mais pour pouvoir bénéficier du supplément, les membres du groupe de traitement devaient retrouver un emploi la première année : s'ils ne pouvaient justifier d'un travail de 120 heures sur une période de 4 semaines la première année, ils perdaient leur droit au supplément de façon définitive. A peu près 98 % des membres du groupe de traitement ont assisté à une orientation décrivant les mécanismes de l'expérimentation. Selon Michalopoulos et al. (2002), le message principal des réunions d'orientation était que le supplément faisait en sorte que « le travail paie ». Les réunions ont également permis d'informer les participants des services disponibles pour les aider à rechercher un emploi. D'après les auteurs, une enquête réalisée ultérieurement par téléphone, a montré que 90 % des participants avaient compris les mécanismes principaux de l'expérience : la condition de 30 heures hebdomadaires travaillées et de reprise de travail la première année pour pouvoir bénéficier de suppléments de revenu généreux. Card et Robbins (1996) soulignent que cette réunion d'information, et d'autres qui ont suivies, faisaient partie du traitement. Elles étaient nécessaires dans la mesure où les incitations financières ne peuvent avoir d'effet que lorsqu'elles sont connues.

9. = (30,000-14250)/2

Résultats

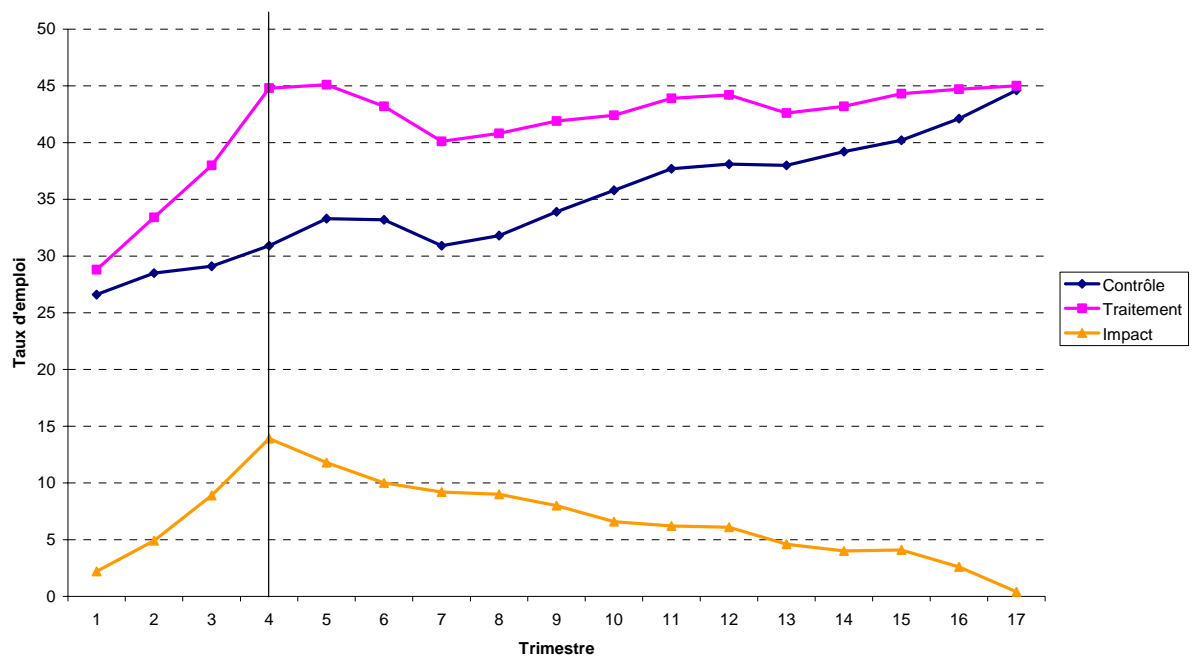
Dans un premier temps, le traitement a eu pour conséquence d'accroître l'emploi des membres du groupe expérimental de manière significative par rapport au groupe de contrôle. On peut observer sur le graphique 1a l'effet du programme sur l'emploi à plein-temps¹⁰. C'est naturellement à la fin de la première année, quand la possibilité de bénéficier de suppléments disparaissait, que l'effet du traitement était le plus important : 30 % des membres du groupe de traitement étaient alors en emploi à plein-temps, soit deux fois plus que les membres du groupe de contrôle (15 %). Selon les auteurs du rapport final sur l'expérimentation (Michalopoulos et *al.*, 2002), cet effet est l'un des impacts les plus importants mesurés lors d'évaluations de programmes d'incitation financière en Amérique du Nord. Au total, 36 % des personnes éligibles au supplément ont trouvé un emploi la première année et ont donc pu bénéficier de versement de SR les trois années suivantes.

L'impact du traitement s'est réduit à partir de la fin de la première année. Au cours des quatre années suivantes, l'emploi à plein-temps du groupe de traitement est resté plutôt stable alors que celui du groupe de contrôle augmente régulièrement. La deuxième année, la différence de taux d'emploi à temps plein entre les deux groupes est en moyenne de 12,6 % ; elle est ensuite égale à 9,3 % la troisième année, 6,1 % la quatrième année. Au milieu de la cinquième année, soit après la fin du traitement, la différence est faible et n'est plus significative. Les membres du groupe de contrôle sont plus souvent employés à temps partiel : les différences en terme de taux d'emploi sont donc plus faibles, en moyenne de 2 points. Au milieu de la cinquième année, les taux d'emploi des deux groupes sont égaux.

La réduction de l'impact du traitement n'est pas surprenante : à la fin de la première année, les personnes n'ayant pas tenu un emploi à plein-temps pendant quatre semaines, soit 64 % du groupe de traitement, n'ont plus de droits spécifiques par rapport aux personnes du groupe de contrôle. Les deux tiers du groupe de traitement sont donc confrontés à partir de la deuxième année aux mêmes incitations que le groupe de contrôle. Ce groupe de « recalés » est constitué des personnes les plus éloignées du marché du travail. Le graphique 1b montre qu'au cours de l'expérience, leur taux d'emploi est inférieur à celui du groupe de contrôle. Mais après la première année les performances en termes d'emploi du groupe de contrôle et du groupe de traitement « recalé » sont similaires. A l'inverse, les personnes ayant réussi à trouver un emploi à plein-temps la première année sont celles ayant les caractéristiques les plus favorables à l'emploi. Il est donc normal que leur taux d'emploi soit nettement supérieur au groupe de contrôle durant toute la période. Par construction, le taux d'emploi à la fin de la première année est très élevé. Partant d'un niveau élevé, il ne peut que décroître. De façon attendue, durant le traitement, la décade est de moins en moins prononcée et le taux d'emploi atteint même un plateau. A la fin du traitement, le taux d'emploi repart à la baisse mais la chute n'est pas brutale (graphique 1b).

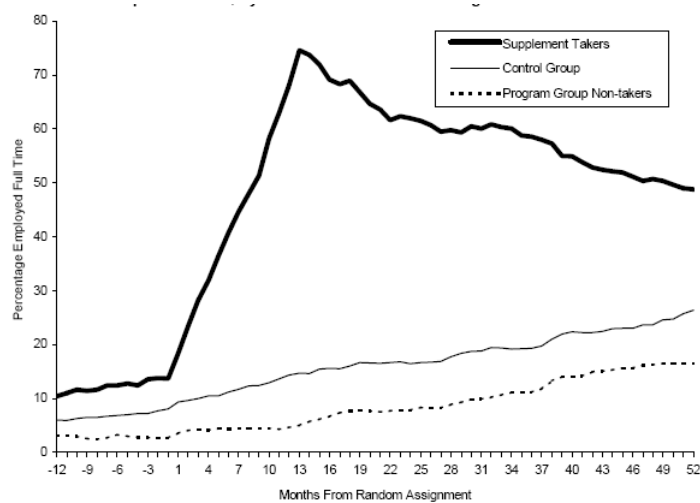
10. Défini par une durée de travail hebdomadaire supérieure à 30 heures au moins une semaine dans le mois.

Graphique 1a : Taux d'emploi trimestriel des groupes de traitement et de contrôle et impact du traitement, PAS – prestataires de longue date.



Source : Michalopoulos, 2005

Graphique 1b : Taux d'emploi trimestriel du groupe éligible au supplément, du groupe des recalés et du groupe de contrôle



Source : Michalopoulos et al., 2002

L'absence d'effet positif significatif du traitement sur l'emploi à l'issue de la période d'observation est troublante. On s'attend à ce que l'effet se réduise mais pas à ce qu'il devienne non significatif. En effet, le groupe de traitement est en moyenne plus en emploi tout le long de l'expérience. Les membres du groupe de traitement acquièrent donc de l'expérience professionnelle à un rythme plus rapide que ceux du groupe de contrôle. Cette expérience professionnelle plus importante devrait leur donner un avantage sur le marché du travail. Or, la différence en termes de salaires horaires à l'issue de l'observation est également très faible et non significative. Comment explique-t-on le fait qu'à l'issue de l'expérimentation, les membres du groupe de traitement en emploi ont travaillé en moyenne 1 an de plus que ceux du groupe de contrôle¹¹ et ne bénéficient pas de salaires significativement supérieurs ? Il y a deux explications possibles : soit (1) l'expérience professionnelle supplémentaire ne procure pas d'avantage significatif ; soit (2) l'expérience supplémentaire procure un avantage mais d'autres éléments viennent réduire cet avantage.

L'expérience professionnelle a-t-elle des effets positifs ? Les évaluateurs essayent de répondre à cette question (Michalopoulos et *al.*, 2002). En analysant la situation des personnes qui étaient en emploi à l'issue de la première année et à l'issue de la quatrième année, ils montrent qu'elles ont bénéficié de hausses de salaires significatives : 41% des membres du groupe de traitement employés aux deux dates et 36% des membres du groupe de contrôle ont vu leur salaire augmenter de plus de 20%. En outre, le taux de croissance des salaires des moins qualifiés augmente aussi vite que celui des plus qualifiés, l'augmentation des salaires ne peut donc pas être uniquement la conséquence de la rapide augmentation du salaire minimum durant la période. Les auteurs concluent que ces résultats confirment l'hypothèse que l'expérience professionnelle accroît les revenus. Cette conclusion est importante car l'idée que le capital humain se déprécie lorsqu'on est éloigné de l'emploi justifie en partie les politiques de lutte contre les trappes à inactivité. En fait, l'augmentation des salaires ne prouvent pas un effet potentiellement bénéfique de l'expérience professionnelle acquise grâce à de plus fortes incitations. Surtout, les auteurs n'avancent pas d'explications quant à l'absence d'effets à long-terme du traitement sur les salaires.

Les personnes éligibles aux suppléments de revenu peuvent être incitées à accepter des emplois moins bien rémunérés. Michalopoulos et *al.* soulignent que les personnes appartenant au groupe de traitement trouvent en moyenne des emplois moins bien rémunérés en taux horaire que celles du groupe de contrôle. Mais comme elles sont également plus nombreuses à travailler et que les personnes supplémentaires sont en moyenne moins qualifiées, on ne peut pas dissocier la part due à un effet de sélection de celle due aux nouvelles incitations. On ne peut pas savoir directement si certaines personnes du groupe de traitement auraient trouvé des emplois mieux rémunérés s'ils ne bénéficiaient pas des suppléments de revenus. C'est pourtant une explication possible de l'absence d'effet sur les revenus du programme : l'expérience professionnelle supplémentaire acquise procurerait un avantage mais d'autre part le supplément de revenu inciterait à écourter la recherche d'emploi et à accepter des emplois plus précaires, moins bien rémunérés ou offrant moins de perspectives de carrière.

Leçons de l'expérimentation

Dans le rapport final sur le projet d'autosuffisance, Michalopoulos et *al.* avancent que l'expérimentation a délivré des « réponses définitives » aux questions auxquelles elle

11. Calcul de l'auteur d'après le tableau 3.1 page 71 (Michalopoulos et *al.*, 2002)

entendait répondre : « Il est clair que les incitations financières comptent dans les décisions d'emplois des bénéficiaires d'allocations sociales. L'incitation à l'emploi à plein-temps contenue dans le programme a accéléré de deux à trois ans le retour à l'emploi des allocataires. L'impact en terme d'emploi est l'un des plus élevés produit lors d'une expérience contrôlée »¹².

Pourtant les résultats sont plus ambigus : Gabel et LeRoy (2003) soulignent que l'objectif principal du programme, l'autosuffisance, n'a pas été atteint. Le fait qu'il n'y ait pas de gains à long terme en termes d'emploi et de salaires devrait préoccuper davantage les évaluateurs. A l'inverse, l'impact en termes d'emploi à l'issue de la première année doit être relativisé. En effet, cet impact est la conséquence d'un mécanisme particulier : à la fin de l'année, les personnes n'ayant pas trouvé d'emploi perdaient l'éligibilité à un revenu potentiel important au cours des trois années suivantes. Ne pas trouver d'emploi pouvait leur faire perdre jusqu'à plusieurs années de salaire en suppléments de revenus. L'expérience montre donc que confrontés à des incitations de cette ampleur, une partie des allocataires de minima sociaux réagissent en reprenant un emploi. Mais les conclusions en termes de politiques publiques sont limitées car il ne paraît ni possible ni souhaitable de généraliser ce type de mécanisme. On peut également souligner que malgré des incitations à l'emploi d'une très grande ampleur la première année, 64% des personnes éligibles n'ont pas pu trouver ou tenir un emploi : si les incitations financières peuvent avoir un impact important sur l'emploi, l'effet positif ne concerne qu'une minorité des personnes sans emploi.

12. *“It is clear that financial incentives do matter to the employment decisions of welfare recipients. The full-time work incentive within SSP accelerated by two to three years welfare recipients' transition to employment. It produced some of the largest employment impacts seen in random assignment program evaluation ».*

Les expériences de compléments de revenus menées par le MDRC

Contexte politique et économique

En 1993 dans son discours sur l'Etat de l'Union, Bill Clinton réitère sa promesse de campagne de « mettre un terme à l'aide sociale telle que nous la connaissons » (« *End welfare as we know it* »). Pendant la campagne électorale de 1992, Clinton avait proposé de limiter les bénéficiaires de l'aide sociale (en fait l'AFDC) dans le temps et d'obliger les allocataires à travailler après deux ans de bénéfice des aides sociales. En 1992, 13,6 millions de personnes bénéficient de l'AFDC, en augmentation de 27 % en 3 ans et de 91 % depuis 1970 (Page et Larner, 1997). L'augmentation du nombre d'allocataires est en partie due à l'expansion de l'AFDC qui bénéficie à toutes les familles pauvres dont l'apporteur principal de ressource est au chômage et non plus seulement aux mères célibataires (ce critère d'éligibilité ayant été critiqué car favorisant les dissolutions maritales). De même que dans les années 70, l'AFDC était critiqué pour être trop désincitatif au travail. Probablement du fait des nombreuses critiques, le montant de la prestation maximale d'AFDC avait reculé de 47 % en termes réels en 20 ans (Miller et al., 2000). De façon parallèle, le taux de pauvreté des familles avec enfants était en constante augmentation.

Dans ce contexte, le MDRC, une organisation spécialisée dans l'expérimentation sociale¹³ propose à un certain nombre d'Etats la mise en place de programmes sociaux expérimentaux à tirage aléatoire. L'organisme met ainsi en place des programmes de *welfare to work*. Certains programmes se concentrent sur les questions d'accompagnement et de formation (voir Gueron et Hamilton, 2002); d'autres expérimentent la limitation dans le temps du bénéfice des aides sociales (voir Bloom, 1999). Trois expériences étudient plus particulièrement la question des incitations au travail : le *Minnesota Family Investment Program* (MFIP), le *Milwaukee's New Hope Project* (*New Hope*), et le *Connecticut Jobs First Program* (*Jobs First*). *Jobs First* combine des éléments d'incitations financières avec une limitation dans le temps des aides sociales. L'objectif principal des expérimentations est d'évaluer en termes d'emploi et de revenus des incitations financières. Un second objectif étant de mesurer les effets de ce type de programme sur le comportement, la santé et la réussite scolaire des enfants¹⁴.

Les expérimentations se sont déroulées dans un contexte économique favorable, notamment celle se déroulant dans le Connecticut à partir de 1996 : durant la période d'expérimentation, le taux de chômage y a baissé de 5,7 % à 2,3 %. Les expérimentations évaluent un impact différentiel sur deux groupes confrontés au même contexte, mais on peut imaginer que l'ampleur de cet impact soit affectée par le contexte économique.

Protocoles mis en place

Le tableau résume les protocoles mis en place dans les trois Etats¹⁵. Il s'agit d'expériences contrôlées avec tirage au sort aléatoire : les membres de l'échantillon sont répartis au hasard dans un groupe de contrôle et dans un ou plusieurs (dans le cas du MFIP) groupes de traitement. Dans les trois Etats, le groupe de contrôle est soumis aux règles de l'AFDC. L'AFDC apporte une prestation essentiellement différentielle aux familles ayant au moins un

13. Organisation fondée en 1974 sous le nom *Manpower Demonstration Research Corporation*.

14. Voir Gennetian et Miller (2000) pour une analyse de l'impact du MFIP sur les enfants.

15. Source : Gennetian et al. (2005) pour le MFIP, Huston et al. (2003) pour *New Hope* et Bloom et al. (2002) pour *Jobs First*.

enfant de moins de 18 ans sous deux conditions : le chef de ménage doit être au chômage ou ne pas travailler plus de cent heures par mois et le revenu disponible de la famille doit être inférieur à un seuil dépendant de la taille de la famille. Dans les ménages de conjoint, le chef de famille doit également avoir travaillé pendant 6 des 13 derniers trimestres.

Les protocoles se différencient sur de nombreux autres points. Le MFIP est l'expérimentation ayant l'échantillon le plus important. C'est également celle dont le protocole expérimental est le plus instructif en termes d'impact des incitations financières à l'emploi. L'expérimentation prévoyait en effet un traitement qui consistait uniquement en de plus fortes incitations financières à l'emploi alors que les deux autres expériences combinaient plusieurs éléments. *New Hope* proposait des emplois d'intérêt général aux membres du groupe de traitement qui souhaitaient travailler mais ne trouvaient d'emplois dans le secteur privé (Huston et al., 2003). *Jobs First* prévoyait une obligation d'accompagnement et de formation en plus des incitations financières et de la limitation dans le temps de l'aide sociale. En théorie, dans *Jobs First*, l'aide sociale était limitée à 21 mois : les personnes n'ayant pas retrouvé d'emploi entre temps se seraient alors retrouvées sans ressource. En pratique, cette règle a fait l'objet de nombreuses dérogations et la suppression des aides sociales étaient laissée à l'appréciation des travailleurs sociaux : si la personne était jugée avoir fait des efforts suffisants de recherche d'emploi, une extension de 6 mois renouvelable était accordée. D'autre part, les familles se retrouvant sans ressources étaient dirigées vers des associations caritatives.

Les trois programmes proposent des suppléments de revenu d'activité différents. Contrairement aux deux autres programmes, *New Hope* conditionne les suppléments à l'emploi à plein temps (au moins 30 heures par semaine). Dans *Jobs First*, les gains sont proportionnels aux revenus : le cumul entre aide sociale et revenus d'activité est intégral jusqu'à un niveau de revenu d'activité égal au seuil de pauvreté. Enfin, dans le MFIP les gains sont dans un premier temps croissants avec le revenu d'activité puis décroissants (à partir de 25h de travail par semaine, l'aide est réduite de 62 centimes par dollar de revenus d'activité supplémentaire), favorisant ainsi les reprises d'emploi à temps partiel : un parent élevant seul deux enfants travaillant 20 heures (respectivement 40) par semaine à \$6 de l'heure gagnait \$237 (respectivement \$148) de plus sous le MFIP que sous l'AFDC.

Le MFIP est la seule expérimentation évaluant l'effet de suppléments de revenus d'activité sur les couples biactifs (Gennetian et al., 2005) : en réduisant l'offre de travail d'un des conjoints, les couples biactifs pouvaient devenir éligibles au supplément de revenus d'activité. Contrairement aux autres expérimentations, l'effet potentiellement désincitatif sur le travailleur supplémentaire de suppléments de revenus d'activité familialisés est ainsi mesuré. Pour les couples le MFIP supprimait deux des conditions d'éligibilité à l'AFDC : la règle des cent heures par mois et celle concernant le travail durant 6 des 13 derniers trimestres ne concernaient pas les membres du groupe de traitement.

Le MFIP débute en 1994 et est généralisé à tout l'État, groupe de contrôle compris, en janvier 1998, mettant ainsi fin à l'expérience contrôlée. A l'inverse, *Jobs First* devient le système en vigueur dans tout l'État du Connecticut dès 1996, mais pour les besoins de l'expérience les foyers du groupe de contrôle sur les sites expérimentaux (New Haven et Manchester) sont les seuls à rester dans le programme de l'AFDC. Ils seront reversés dans le programme *Jobs First* en 2001, après la période de suivi de l'expérimentation (Bloom et al., 2002).

Tableau 1 : Protocoles mis en place

	MFIP	New Hope	Jobs First
Sites	7 contés dans le Milwaukee	2 quartiers pauvres de Milwaukee	New Haven et Manchester, Connecticut
Critères de sélection de l'échantillon	Parents inscrits ou s'inscrivant au bureau d'aide sociale 2 échantillons : - foyers ayant reçu aide sociale (afdc) pendant 24 des 36 derniers mois. - inscrits récents	Tous les adultes suivant deux critères : (1) revenus inférieurs à 200% du taux de pauvreté (2) volonté de travailler au moins 30 heures par semaine	Foyers éligibles à l'AFDC
Nombre de foyers dans l'échantillon	11,473 foyers 9,217 célibataires et 2,256 foyers de conjoints	1,300 foyers	5,000 foyers
Traitements	2 traitements : (1) Supplément de revenu d'activité (MFIP incitations seules) (2) Supplément de revenu et obligation d'accompagnement (MFIP)	4 éléments dans le traitement : (1) Supplément de revenu d'activité pour temps plein (30h/sem) (2) Subvention pour assurance maladie (3) Subvention pour garde d'enfant (4) Accès à des emplois d'intérêt général	3 éléments dans le traitement : (1) Cumul intégral de l'aide sociale et des revenus d'activité jusqu'au seuil de pauvreté (2) Obligation d'accompagnement (3) Limite à 21 mois de la perception du revenu minimum garanti (avec dérogation possible)
Durée de l'expérimentation	4 ans	3 ans	4 ans
Début de l'expérimentation	1994	1994	1996

Résultats

Les graphiques 2a à 2e présentent les résultats en termes d'emploi des trois expériences. Les lignes verticales représentent la fin du traitement.

Dans le MFIP, l'échantillon était divisé en deux groupes : un groupe composé de personnes ayant reçu l'AFDC pendant 24 des 36 mois précédents et un groupe composé des inscrits récents et des personnes s'inscrivant à l'aide sociale. Les résultats en termes d'emploi sont présentés séparément (graphiques 2a et 2b). L'expérimentation permet également de représenter de façon séparée l'impact du traitement « incitations financières seules » et de celui « incitations financières et obligation d'accompagnement ». En ce qui concerne les allocataires de long terme, la différence de taux d'emploi est élevée pendant la première année d'expérimentation. A la fin de la première année, le taux d'emploi du groupe de traitement est supérieur de 12,9 % à celui du groupe de contrôle. La moitié de cet impact (6,1 %) peut être attribué aux incitations financières. C'est au troisième trimestre que l'impact des incitations financière est le plus élevé (+ 8,2 %). L'impact décroît dans un deuxième temps et s'annule au 7^{ème} trimestre (probablement parce que les personnes qui sont retournées le plus vite sur le marché du travail ont trouvé des emplois inadaptés). A la fin de la troisième année, l'impact des incitations financières est positif (+ 4,8 %) et significatif. A la fin du traitement, l'impact est positif (+ 2,4 %) mais non significatif. Il reste non significatif pendant la période d'observation suivant le traitement. A la fin de l'observation, l'impact est nul : le groupe de contrôle et les deux groupes de traitement ont le même taux d'emploi. De même que les incitations financière, l'effet propre de l'accompagnement est élevé durant les premières années avant de décroître et devenir non significatif à l'issue du traitement. En ce qui concerne les allocataires récents, les incitations financières ont un impact non significatif tout au long de l'observation (graphique 2b). L'accompagnement a un impact positif (+ 3,7 %) durant l'expérimentation, mais l'effet décroît et s'annule à l'issue de l'observation. L'accompagnement et les incitations financières ont ici pour effet d'accélérer le retour à l'emploi des personnes concernées mais n'ont pas d'effet à long terme sur le capital humain : à l'issue de la quatrième année, l'impact du traitement sur les salaires est non significatif sur les deux échantillons.

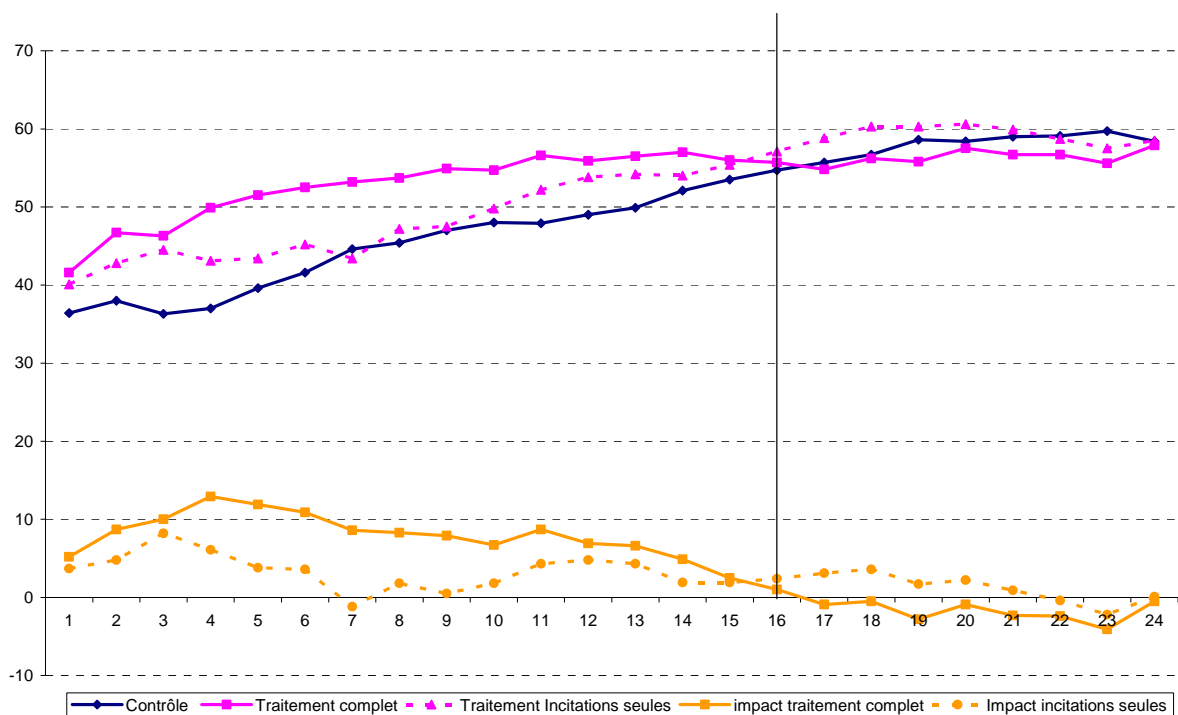
Le MFIP permet de suivre l'emploi parmi les couples de conjoints. Le graphique c représente la proportion de couples où les deux conjoints sont en emploi pour les groupes de contrôle et de traitement pendant les 6 années suivant le début de l'expérience. L'impact initial est négatif : le MFIP réduit l'emploi du travailleur secondaire, le plus souvent la mère, parmi les couples avec enfants (Michalopoulos, 2005). L'effet négatif décroît avant la fin de l'expérimentation, il est non significatif à l'issue de l'expérience. Il est possible que cette décreue soit due à une anticipation de la fin de l'expérimentation. Il est également possible que l'impact négatif initial soit la conséquence d'un arbitrage inter-temporel où un retrait du marché du travail est anticipé pour profiter de conditions favorables temporaires. Ceci illustre une des limites des expérimentations sociales : on ne connaît ni le sens ni l'ampleur du biais de durée limitée.

L'impact de *New Hope* sur l'emploi est globalement positif (graphique 2d). Mais comme les différents éléments du traitement ont été donnés à tous les membres, on ne peut pas distinguer l'effet de l'incitation financière et celui de l'accompagnement. On ne peut pas non plus distinguer les effets d'offre des effets de demande de travail étant donné qu'un élément du traitement était un accès facilité à des emplois publics d'intérêt général. De même que dans les autres expérimentations, l'impact du traitement sur l'emploi et le revenu d'activité est nul :

les mois ou les années d'expérience professionnelle supplémentaire des membres du groupe de traitement n'ont pas d'impact à long terme. Le traitement accélère le retour à l'emploi mais ne donne pas d'avantage particulier à ces bénéficiaires.

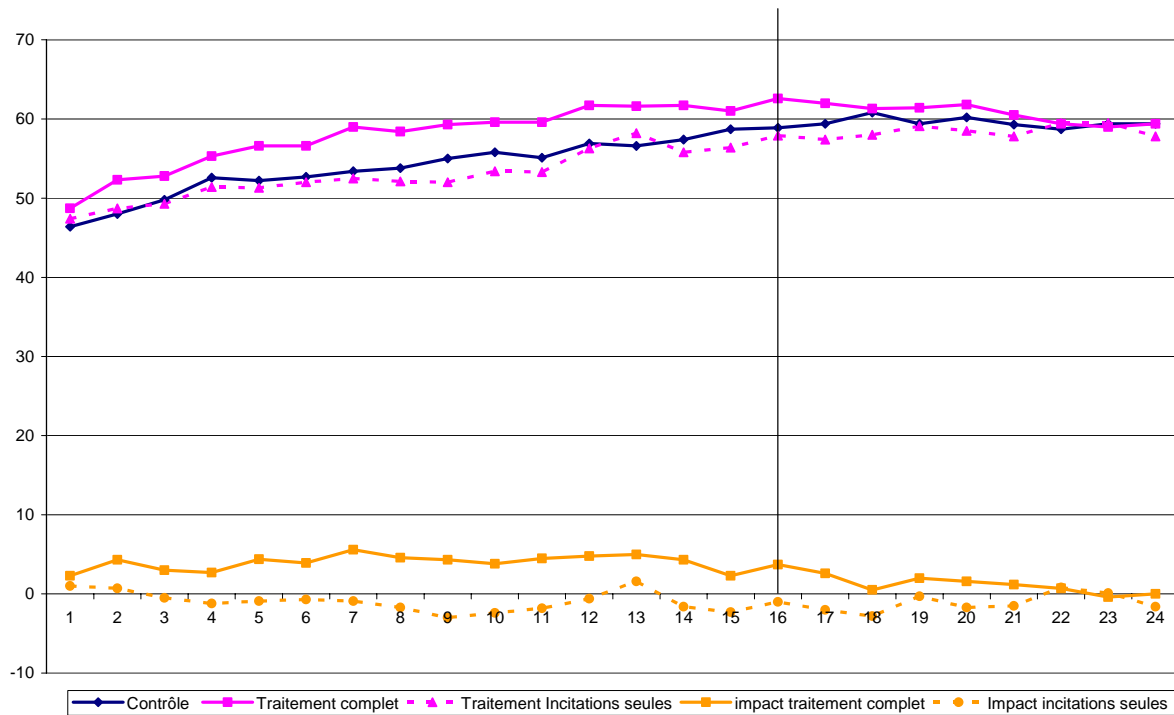
L'impact de *Jobs First* sur les taux d'emploi est également globalement positif (graphique 2e). De même que pour *New Hope*, l'effet des incitations financières ne peut être distingué de celui de l'obligation d'accompagnement. De plus, à l'intérieur des incitations financières, on ne peut distinguer l'effet des suppléments de revenu d'activité de celui de la suppression du revenu minimum garanti au-delà de 21 mois. Les effets à long terme de l'expérimentation ne sont pas connus : contrairement aux autres expérimentations, un suivi après la période de traitement n'était pas prévu.

Graphique 2a: Taux d'emploi trimestriel des groupes de traitement et de contrôle et impact des traitements, MFIP allocataires de long terme.



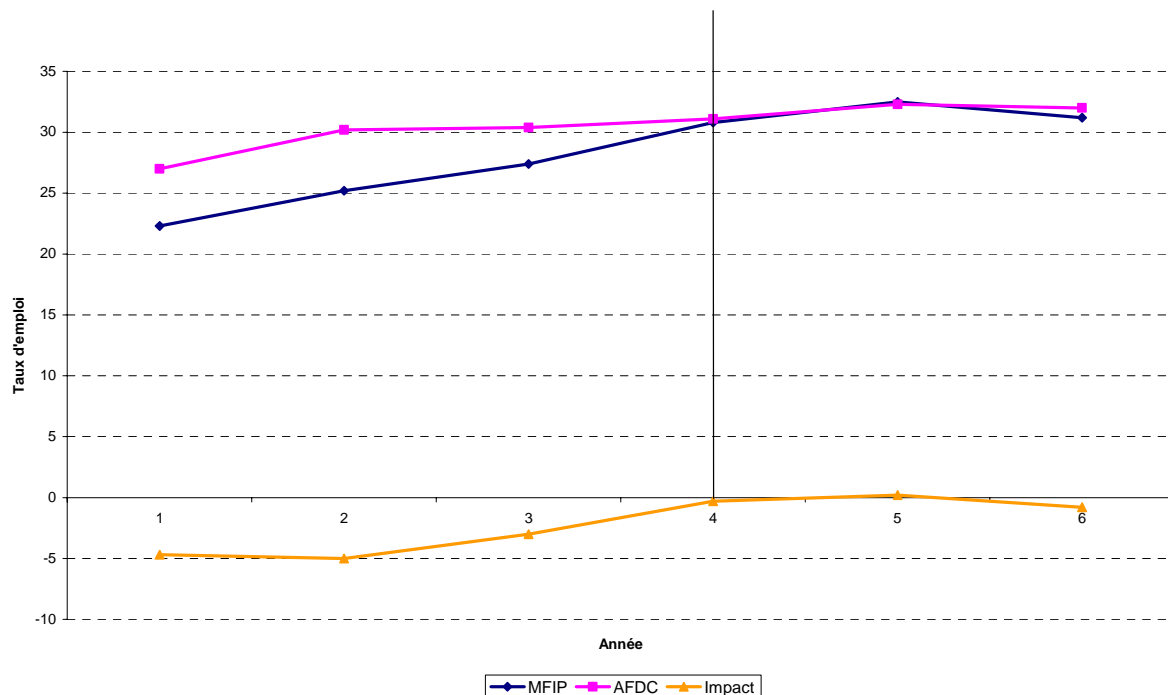
Source : Michalopoulos

Graphique 2b: Taux d'emploi trimestriel des groupes de traitement et de contrôle et impact des traitements, MFIP allocataires récents.

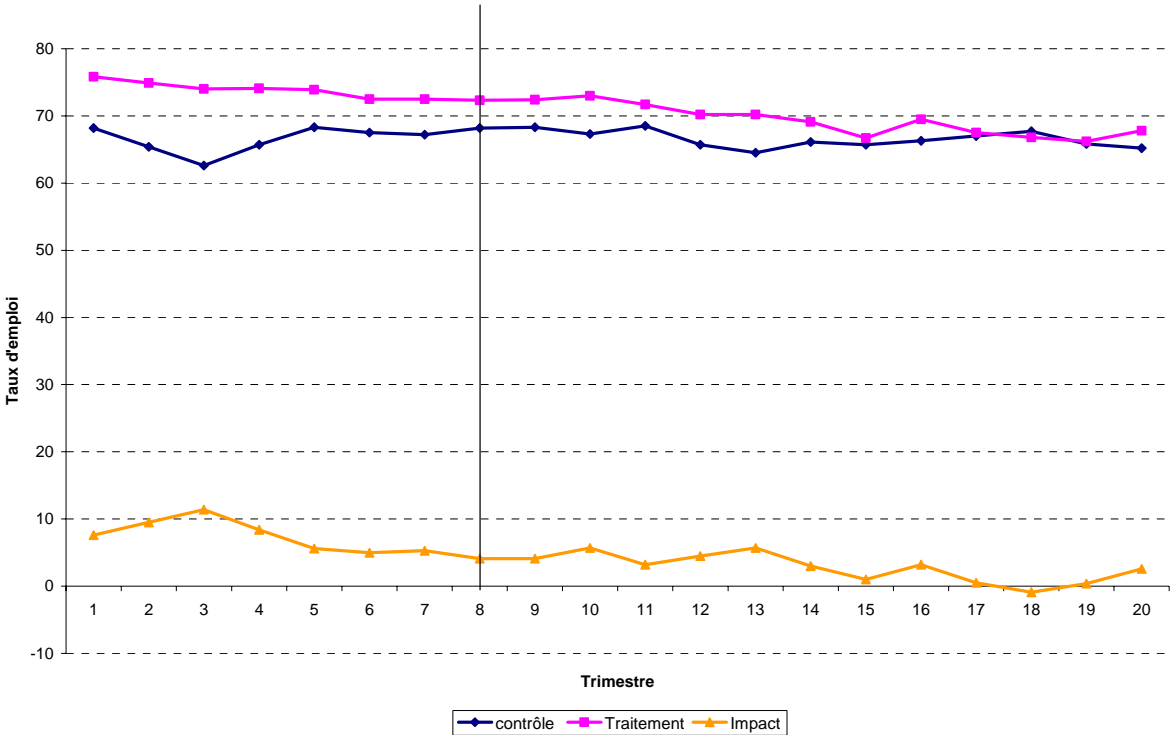


Source : Michalopoulos

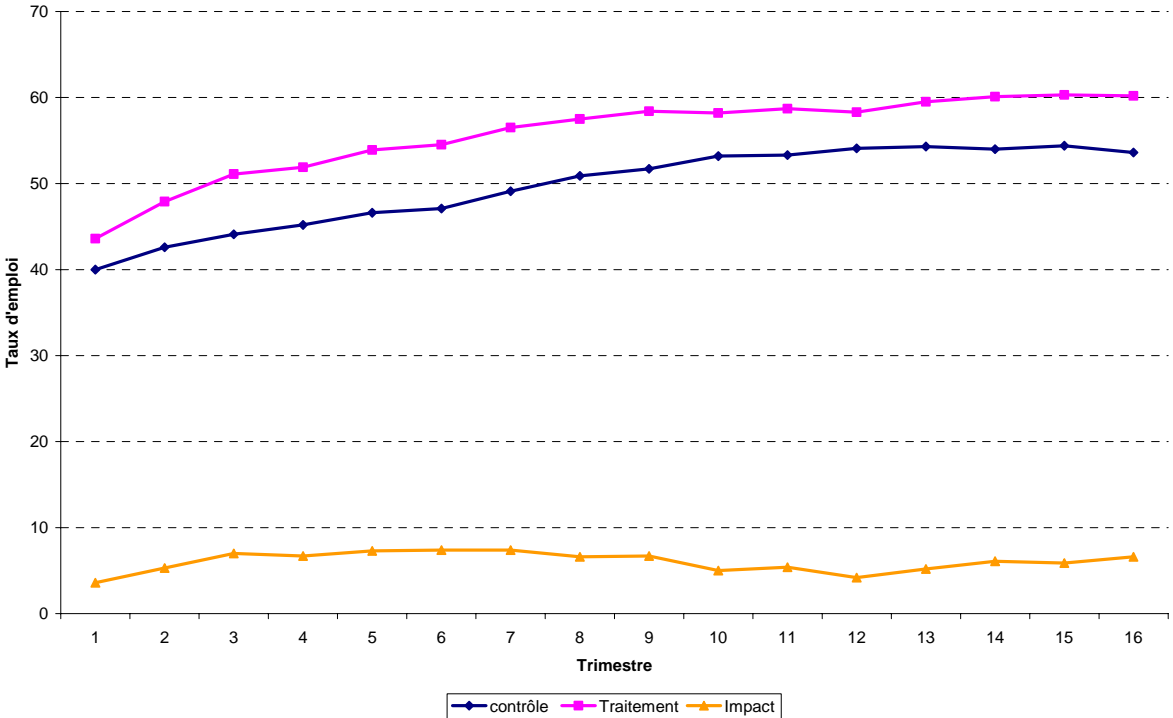
Graphique 2c: Proportion de couples où les deux conjoints sont en emploi, groupe de contrôle, groupe de traitement et impact, MFIP.



Graphique 2d: Taux d'emploi trimestriel des groupes de traitement et de contrôle et impact du traitement, New Hope.



Graphique 2e: Taux d'emploi trimestriel des groupes de traitement et de contrôle et impact du traitement, Jobs First.



Leçons des expérimentations

L'intérêt du MFIP par rapports aux expérimentations précédentes est de donner une vision plus complète de l'effet des incitations financières puisque l'impact sur les couples en emploi est également mesuré. L'expérience contrôlée confirme des résultats que l'on retrouve ailleurs dans la littérature (voir Eissa et Hoynes, 1998) : les aides familialisées peuvent avoir un effet négatif sur l'emploi du travailleur secondaire.

De même que dans le projet d'autosuffisance, dans l'expérience du MFIP, l'impact en termes d'emploi des incitations financières décroît pendant l'expérimentation dès la seconde année, mais contrairement au projet d'autosuffisance, le traitement est constant pendant l'expérimentation. L'impact à court-terme peut être biaisé pour un certains nombres de raison. Il est donc intéressant de conduire ce type d'expérimentation pendant quelques années afin que l'impact du traitement soit stabilisé. Ici, on ne sait pas quel aurait été l'impact différentiel du traitement si l'expérience contrôlée avait duré plus longtemps. Une limite de l'expérimentation sociale apparaît : la réduction de l'impact du traitement au cours de l'expérimentation est difficile à interpréter : est-ce l'effet d'un biais de durée limité ou d'effets d'apprentissage ? On ne peut pas vraiment savoir si la mise en place d'incitations financières ne fait qu'accélérer les reprises d'emploi ou si ces politiques élèvent durablement les taux d'emploi. Le fait que les effets des programmes d'incitation disparaissent rapidement après la fin de l'expérimentation est un argument en faveur de la première explication.

Conclusion : quelles leçons pour l'expérimentation du revenu de solidarité active ?

Les résultats des expérimentations nord-américaines sont difficiles à généraliser : ils sont le produit de protocoles et de contextes spécifiques. Ils tendent néanmoins à montrer que les incitations financières permettent d'accélérer la reprise d'emploi des allocataires de minima sociaux. Les impacts sont généralement faibles et non significatifs chez les hommes et les allocataires de long terme et plus élevés chez les femmes seules avec enfants et les allocataires récents. Conformément aux prédictions théoriques et aux résultats non expérimentaux, les programmes familialisés ont un impact négatif sur l'emploi des femmes en couple avec enfants. L'expérimentation du Revenu de Solidarité Active ne concerne que les allocataires de minima sociaux (Rmi, Api), le résultat attendu est donc un impact positif sur l'emploi. L'expérimentation permet d'évaluer l'ampleur de cet effet.

Les expérimentations nord-américaines montrent que les effets positifs des compléments de revenu ne sont pas durables : quelques mois après la fin de l'expérimentation, il n'y a plus de différence d'emploi entre les groupes expérimentaux et les groupes de contrôle. Il n'y a pas de gains liés à l'accumulation de capital humain lors des périodes d'emploi : les emplois repris seraient de faible qualité, offrant peu de perspectives de carrière. En ce qui concerne les protocoles expérimentaux, il est nécessaire que les traitements soient suffisamment long afin de mesurer correctement les impacts en termes d'emploi : les effets potentiellement négatifs dus à la mauvaise qualité des emplois repris n'apparaissent qu'à moyen terme. Le suivi des groupes de traitement et de contrôle après la période de traitement doit permettre d'évaluer les effets à long terme de l'expérimentation. Le biais de durée limitée peut être testé en évaluant des traitements de durée différente.

Concernant les autres aspects des protocoles expérimentaux, plusieurs leçons peuvent être tirées des expériences nord-américaines. Il est souhaitable dans les expérimentations qui combinent des éléments d'incitation financières et des éléments d'accompagnement de créer plusieurs groupes de traitement afin de pouvoir distinguer l'effet propre de ces éléments. Les expérimentations américaines d'impôt négatif ont été critiquées car elles ne pouvaient répondre à la question de l'effet d'une généralisation d'un tel programme, l'introduction n'étant testée que sur des populations spécifiques (en termes de communauté et/ou de revenu). Il apparaît important que l'échantillon de l'expérimentation soit représentatif de la population potentiellement concernée par la généralisation de la réforme. Dans le cadre de l'expérimentation du RSA, l'impact potentiellement négatif de l'instrument sur l'offre de travail des couples biactifs n'est pas évalué puisque seuls les allocataires de minima sociaux sont éligibles. Du point de vue des politiques publiques, on ne pourra donc pas répondre à certaines interrogations importantes sur la mise en place du RSA : quel effet sur l'emploi total et sur le nombre total d'heures travaillées ? Quel serait le coût *ex post* d'une telle mesure ?

Contrairement aux expérimentations nord-américaines, l'expérimentation du RSA n'utilise pas le tirage aléatoire pour former les groupes de traitement et de contrôle. Les départements choisissent des territoires de traitement à l'intérieur desquels tous les allocataires du Rmi et de l'Api constituent les groupes de traitement (un par département). Les groupes de contrôle sont constitués des allocataires du Rmi et de l'API sur des territoires de contrôle sélectionnés par les évaluateurs. L'utilisation de territoires d'expérimentation et de contrôle a pour avantage, par rapport au tirage au sort d'individus dans un même territoire, de réduire le problème lié aux effets d'éviction entre allocataires de minima sociaux si ces territoires correspondent à

des bassins d'emploi relativement étanches¹⁶. Cette méthode de contrôle permet également d'éviter les problèmes éthiques et administratifs liés au tirage aléatoire des groupes expérimentaux. D'un point de vue statistique, cette méthode n'est pas aussi adéquate qu'une réelle expérience contrôlée mais pourraient constituer un compromis satisfaisant si les territoires de contrôle ont des caractéristiques proches aux territoires de traitement. Il convient néanmoins d'analyser la sélection des territoires de traitement par les départements afin de repérer d'éventuels biais de sélection qui seraient préjudiciables à l'expérimentation.

16. Puisque tous les Rmistes et les Apistes sur les territoires expérimentaux bénéficient du traitement. Les autres catégories de la population peuvent néanmoins également être victimes d'effets d'éviction.

Références bibliographiques

Aaron H. 1975 : « Cautionary notes on the experiment » in *Work Incentives and Income Guarantees: the New Jersey Negative Income Tax Experiment*, Brookings Institution.

Ashenfelter O., D. Ashmore et O. Deschênes, 1999 : « Do Unemployment Insurance Recipients Actively Seek Work? Randomized Trials in Four U.S. States », *NBER Working Paper Series*, n°6982

Bloom D., S. Scrivener, C. Michalopoulos, P. Morris, R. Hendra, D. Adams-Ciardullo, J. Walter, 2002 : *Jobs First, Final Report on Connecticut's Welfare Reform Initiative*, MDRC.

Burtless G., 1995 : «The Case for Randomized Field Trials in Economic and Policy Research», *Industrial and Labor Relations Review*, vol. 39, n°1.

Burtless G et L Orr, 1986 : «Are Classical Experiments Needed for Manpower Policy? », *The Journal of Human Resources*, Vol. 21, n°4.

Burtless G et D. Greenberg, 1982 : «Inferences Concerning Labor Supply Behavior Based on Limited-Duration Experiments», *The American Economic Review*, Vol. 72, No. 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

Card D. et D. Hyslop, 2005 : « Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers », *SRDC Working Paper Series*, n°05-02

Coyle D. et A. Wildavsky, 1986 : « Social Experimentation in the Face of Formidable Fables » in Munnell, *Lessons from the Income Maintenance Experiments*, Federal Reserve Bank of Boston

Fisher R.A., 1966 : *The design of Experiments*, Oliver and Boyd.

Fougère D., 2000 : « Expérimenter pour évaluer les politiques d'aide à l'emploi : les exemples anglo-saxons et nord-européens », *Revue Française des Affaires sociales*, n°1, janvier-mars.

Friedman, M., 1962 : *Capitalism and Freedom*. University of Chicago Press.

Gennetian L., C. Miller et J. Smith, 2005 : *Turning Welfare into a Work Support, Six-Year Impacts on Parents and Children from the Minnesota Family Investment Program*, MDRC, New York.

Gennetian L. et C. Miller, 2000 : *Reforming Welfare and Rewarding Work: Final Report on the Minnesota Family Investment Program, Volume 2: Effects on Children*, MDRC, New York

Heckman J., 1992 : « Randomization and Social Policy Evaluation », dans *Evaluating Welfare and Training Programs*, Harvard University Press

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. La nouvelle équation sociale*, Rapport de la Commission Familles, vulnérabilité, pauvreté. Ministère des solidarités, de la santé et de la famille, La Documentation Française, Paris.

Hum D. et W. Simpson, 1993 : « Economic Response to a Guaranteed Annual Income : Experience from Canada and The United States », *Journal of Labor Economics*, Vol. 11, n°1

Huston A., C. Miller, L. Richburg-Hayes, G. Duncan, C. Eldred, T. Weisner, E. Lowe, V. McLoyd, D. Crosby, M. Ripke, C. Redcross, 2003: *New Hope for Families and Children : Five-year Results of a Program to Reduce Poverty and Reform Welfare*, MDRC.

Keeley M. et P. Robins, 1980 : « Experimental Design, the Conlisk-Watts Assignment Model, and the Proper Estimation of Behavioral Response », *The Journal of Human Resources*, Vol. 15, n°4.

Levine R., 1975 : « How and Why the Experiment Came About » in *Work Incentives and Income Guarantees: the New Jersey Negative Income Tax Experiment*, Brookings Institution.

Mahoney B. et M. Mahoney, 1975: «Policy Implications: A sceptical View » in *Work Incentives and Income Guarantees: the New Jersey Negative Income Tax Experiment*, Brookings Institution.

Metcalf C., 1974 : «Predicting the Effects of Permanent Programs from a Limited Duration Experiment », *The Journal of Human Resources*, Vol. 9, n°4

Michalopoulos C., D. Tattre, C. Miller, P.K. Robins, P Morris, D. Gyarmati, C. Redcross, K. Foley, R. Ford, 2002 : *Making Work Pay, Final Report on the Self-Sufficiency Project for Long-Term Welfare Recipients*, Social Research and Demonstration Corporation.

Michalopoulos C., 2005 : *Does Making Work Still Pay? An Update on the Effects of Four Earnings Supplement Programs on Employment, Earnings, and Income*, MDRC.

Mijanovich T., D. Long., 1995: *Creating an Alternative to Welfare : First-Year Findings on the Implementation, Welfare Impacts, and Costs of the Self-Sufficiency Project*, Social Research and Demonstration Corporation.

Miller C., V. Knox, L. Gennetian, M. Doodoo, J.A. Hunter, C. Redcross, 2000 : *Reforming Welfare and Rewarding Work : Final Report on the Minnesota Family Investment Program, Volume 1 : Effects on Adults*, MDRC, New York

Moffit R., 2003 : «The Negative Income Tax and the Evolution of U.S. Welfare Policy», *Journal of Economic Perspectives*, Vol. 17, n°3

Munnell A., 1986 : *Lessons from the Income Maintenance Experiments*, Proceedings of a Conference Held at Melvin Village, Federal Reserve Bank of Boston, Conference Series n°30

Perez C., 2000 : « L'évaluation expérimentale des programmes d'emploi et de formation aux Etats-Unis : éléments de critique interne », *Revue Française des Affaires sociales*, n°1, janvier-mars.

Rees A. et H. Watts, 1975 : « An Overview of the Labor Supply Results », in *Work Incentives and Income Guarantees: the New Jersey Negative Income Tax Experiment*, Brookings Institution.

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.

Rossi P. et K Lyall, 1976: *Reforming Public Welfare. A Critique of the Negative Income Tax Experiment*, Russel Sage Foundation, New York

Skidmore F., 1975 : «Operational Design of the Experiment », in *Work Incentives and Income Guarantees: the New Jersey Negative Income Tax Experiment*, Brookings Institution.

Whiteford P., 1981 : « Work Incentive Experiments in the United States and Canada », Department of Social Security, Research Paper n°12.

Widerquist K., 2005 : « A failure to communicate: what (if anything can we learn from the negative income tax experiments? », *Journal of Socio-Economics*, 34.