

ÉVALUATION DE LA QUALITÉ DES DONNÉES D'ENQUÊTE PAR APPARIEMENT D'ENREGISTREMENTS DU RECENSEMENT ET D'UNE ENQUÊTE

Stephanie Freeth, Amanda White et Jean Martin¹

RÉSUMÉ

À la suite des trois derniers recensements en Grande-Bretagne, on a étudié la non-réponse à de grandes enquêtes gouvernementales auprès des ménages par couplage des adresses échantillonnées aux fins d'enquêtes ayant eu lieu à l'époque de ces recensements et des enregistrements individuels du recensement pour ces mêmes adresses. Dans le présent document, nous décrivons sommairement le plan de conception de l'étude de non-réponse d'enquête en couplage avec le recensement britannique de 2001. Il y est question de 10 enquêtes dont la conception et les taux de réponse varient amplement. La grande caractéristique de cette étude est le vaste recours à des données auxiliaires et à une modélisation multiniveaux en vue de la constatation des effets aux niveaux de l'intervieweur, du ménage et de la région.

MOTS CLÉS : Non-réponse; Effets d'intervieweur; Effets de ménage; Effets de région

1. INTRODUCTION

L'Office for National Statistics (ONS) est chargé de la réalisation du recensement décennal de la population en Angleterre et au pays de Galles. À la suite des recensements de 1971, 1981 et 1991, la division des enquêtes sociales (Social Survey Division ou SSD) de cet organisme, qui mène nombre des grandes enquêtes gouvernementales auprès des ménages, a étudié la non-réponse à d'importantes enquêtes de cette nature. Dans ces études de non-réponse d'enquête en couplage avec le recensement, on a apparié les adresses échantillonnées aux fins d'enquêtes ayant eu lieu à l'époque des recensements et les enregistrements individuels du recensement pour ces mêmes adresses. Comme on dispose de toutes les variables de recensement pour les adresses répondantes et non-répondantes, c'est là un très puissant moyen d'examen des caractéristiques des non-répondants, de mesure du biais de non-réponse et d'évaluation des méthodes de correction de biais.

Pour les études qui se sont faites après les recensements de 1971 et 1981, les ensembles de données effectifs de couplage n'étaient pas à la disposition des méthodologistes appelés à analyser la non-réponse, mais seulement certains tableaux de valeurs agrégées, ce qui a limité l'analyse à des comparaisons descriptives des caractéristiques des répondants et des non-répondants et à une mesure de biais pour les caractéristiques du recensement. En 1991 cependant, les méthodologistes ont pu disposer des micro-enregistrements de couplage dans de strictes conditions convenues de confidentialité, d'où la possibilité d'y aller d'une modélisation statistique bien plus fine en vue de l'étude des liens entre les variables liées au biais de non-réponse et d'une évaluation des diverses méthodes de pondération de données en compensation de biais (figure 1). L'étude de 1991 a porté sur cinq enquêtes, et une comparaison de résultats d'enquêtes d'une conception très variable a donc été possible. Les principaux résultats en sont présentés dans Foster (1998).

¹ Stephanie Freeth, Amanda White et Jean Martin, Office for National Statistics, Royaume-Uni, 1, Drummond Gate, Londres, Royaume-Uni, SW1V 2QQ.

En 2001, le recensement a eu lieu au Royaume-Uni en avril et, en 2000, l'ONS a commencé à travailler à une étude de la non-réponse d'enquête en couplage avec le recensement de 2001 (étude CNR). Dans le présent document, nous décrivons sommairement la conception de l'étude CNR et exposons les progrès du projet à ce jour.

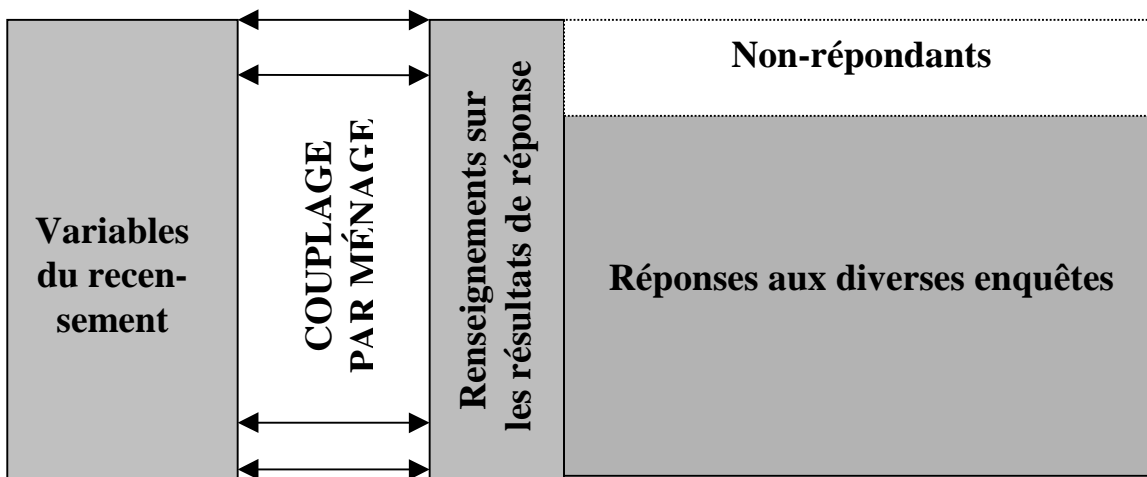


Figure 1 Conception de l'ensemble de données de l'étude de la non-réponse d'enquête en couplage avec le recensement de 1991

2. CONCEPTION DE L'ÉTUDE DE LA NON-RÉPONSE D'ENQUÊTE EN COUPLAGE AVEC LE RECENSEMENT DE 2001

Dans son champ d'observation, l'étude CNR de 2001 est plus vaste que les précédentes. Elle porte sur dix enquêtes dont les plans de conception et les taux de réponse (qui oscillent actuellement entre un peu moins de 60 % et plus de 80 %) sont variables. Elle s'appuie sur les études antérieures de même nature qui ont été réalisées par l'ONS, ainsi que sur des recherches menées aux États-Unis par Groves et Couper (1998), qui ont proposé des modèles conceptuels des facteurs qui déterminent les probabilités de communication de l'intervieweur avec le ménage échantillonné et de consentement à l'interview après la prise de contact.

Ceux-ci énumèrent quatre grandes catégories de facteurs :

- caractéristiques de la région;
- caractéristiques du ménage;
- caractéristiques de conception de l'enquête;
- caractéristiques de l'intervieweur.

Ensemble, ces facteurs influent sur les probabilités de communication, l'interaction du ménage et de l'intervieweur et, par conséquent, les probabilités de collaboration après la prise de contact. Groves et Couper soulignent que ce ne sont pas en soi les caractéristiques démographiques qui déterminent la non-réponse, mais que les gens ayant certaines caractéristiques sont susceptibles d'avoir un mode de vie ou des attitudes qui déterminent la facilité de les joindre à leur domicile ou de les persuader de participer à une enquête. Les intervieweurs ont une énorme incidence sur les résultats de réponse. On a démontré que les attitudes et les stratégies qui sont les leurs dans leur travail et les traits de leur comportement en présence d'un ménage tiennent une grande place dans ces résultats. L'étude CNR vise à mesurer le plus de facteurs possible qui influent vraisemblablement sur la réponse en vue d'un examen de l'interaction des caractéristiques de la région, du ménage, de l'intervieweur et de la conception d'enquête dans leur effet sur la non-réponse.

3. DONNÉES À UTILISER

3.1 Aperçu des données

L'information à exploiter dans l'étude CNR comprend les enregistrements du recensement et les données sur les adresses ou les ménages échantillonnés aux fins d'enquêtes réalisées à l'époque du Recensement de 2001 (qui a précisément eu lieu le 29 avril 2001). Il s'agit d'extraire les données de recensement pour les ménages échantillonnés dans les enquêtes étudiées et de les apparier avec les données d'enquête correspondantes. L'étude 2001 se distingue par l'usage prévu d'une grande quantité de données auxiliaires en complément des données de recensement disponibles. Cette information complémentaire consiste principalement en données d'observation recueillies par les intervieweurs sur le terrain aux adresses sélectionnées. La base de données de l'étude renfermera les renseignements suivants :

- informations concernant les régions échantillonnées;
- informations concernant les ménages échantillonnés;
- caractéristiques du plan d'enquête;
- informations à propos des intervieweurs;
- information sur le comportement des intervieweurs et résultat des visites à chaque adresse.

3.2 Informations concernant les régions

Le gros des renseignements dont nous disposerons sur les régions échantillonnées pour l'étude viendra du recensement, mais avec des données puisées à d'autres sources. Le choix de variables n'est pas encore définitif, mais il est probable que l'on retienne les variables suivantes :

- densité de population;
- milieu urbain ou rural;
- taux de criminalité;
- taux de chômage;
- proportion de propriétaires-occupants;
- proportion de logements multifamiliaux.

Nous utiliserons aussi les données consignées par les intervieweurs sur les adresses échantillonnées, ce qui pourra nous aider à prévoir la facilité de communication ou de mise à contribution. Il peut s'agir, par exemple, du bon état des bâtiments d'une région, que celle-ci soit surtout à vocation résidentielle ou commerciale. Nous étudierons en outre divers classements régionaux (ACORN, MOSIAC, etc.) actuellement disponibles. Les données des régions peuvent être reliées aux échantillons de toute enquête générale auprès de la population que nous réalisons et pourront donc servir à des corrections courantes de non-réponse (ainsi qu'à l'analyse et à la stratification d'échantillons). L'étude CNR nous permettra d'évaluer à quel point les données des régions peuvent compenser pour le biais de non-réponse et de quelle façon on peut le mieux les exploiter.

3.3 Informations concernant les ménages

L'ensemble de données de l'étude comportera toutes les variables du recensement. Il y aura aussi les observations des intervieweurs sur les caractéristiques des ménages sélectionnés, ce qui pourrait nous aider à prévoir la facilité de communication (présence d'un téléphone d'accès ou d'autres obstacles à la prise de contact, par exemple) ou de mise à contribution.

3.4 Caractéristiques du plan d'enquête

Toutes les enquêtes visées par l'étude se font par interviews sur place, mais différent pour ce qui est des caractéristiques de conception suivantes :

- renseignements obtenus pour tous les membres du ménage ou pour un adulte choisi;

- interview en personne de tous les adultes ou information par procuration permise;
- règles de réponse qui déterminent le taux de réponse;
- durée moyenne de l'interview;
- durée des travaux de terrain;
- sujets traités par l'enquête;
- collecte de données hors interview (journaux et relevés, par exemple).

Par les indications de conception d'enquête qu'il contiendra, l'ensemble de données permettra une analyse des effets de conception des enquêtes visées.

3.5 Caractéristiques des intervieweurs

On appréhendera non seulement les régions et les ménages échantillonnés et les caractéristiques de conception d'enquête, mais aussi les caractéristiques des intervieweurs. Ces dernières indications viennent d'une étude des intervieweurs réalisée en juillet 2001 par l'ONS. Cette enquête ressemble à une enquête réalisée en 1998 dans le cadre d'un projet international (Martin et Beerten, 1999; Hox, 1999). Nous avons recueilli les données suivantes sur les intervieweurs :

- caractéristiques sociodémographiques;
- durée et nature de l'expérience des enquêtes;
- rendement (par les taux de réponse et d'autres facteurs pendant une année);
- confiance dans la capacité de faire répondre les ménages;
- connaissance des techniques qui favorisent la réponse;
- stratégies employées pour amener les gens à répondre.

3.6 Comportement des intervieweurs et résultat des visites

Comme dans les études précédentes, il y aura des renseignements sur les visites des intervieweurs aux diverses adresses sélectionnées et sur le résultat de chacune de ces visites. D'autres données sans doute utiles porteront sur l'interaction entre l'intervieweur et l'enquêté lors de la présentation initiale de l'enquête à domicile. Ce sont là des indications de « pas de porte » que consignent les intervieweurs à l'issue de chaque visite. Nous avons ainsi recueilli les renseignements suivants :

- heure et jour de chaque visite;
- résultat de chaque visite;
- nombre de visites pour la prise de contact;
- nombre de visites pour l'interview complète après la prise de contact;
- interaction de pas de porte de l'intervieweur et de l'enquêté avec notamment les aspects suivants :
 - questions et remarques de l'enquêté;
 - motifs cités pour ne pas consentir à l'interview.

4. AUTRES CARACTÉRISTIQUES DE CONCEPTION

Il y aura couplage recensement-enquête pour environ 5 000 adresses dans chaque enquête visée. Nous prévoyons une foule d'analyses pour chaque enquête et un échantillon de 5 000 adresses nous donnera assez de ménages non répondants pour des analyses par enquête. Comme par le passé, les ménages répondants ne seront pas sous-échantillonnés, car cela nuirait aux modèles statistiques à élaborer. Nous inclurons des adresses en Écosse et au pays de Galles pour les enquêtes qui visent ces pays. Idéalement, nous voulons mettre en couplage d'enregistrements les adresses où la prise de contact a lieu le plus près possible du jour du recensement (29 avril 2001). C'est ce qu'on a fait les années précédentes en prenant des adresses sélectionnées pour interview d'enquête dans les mois qui précèdent et suivent le jour du recensement. En 2001, les données de recensement seront appariées avec les données d'enquête pour des adresses en sélection d'interview depuis avril. Il est impossible de prendre des adresses antérieures à avril à cause des

problèmes de compatibilité de données causés par l'adoption ce même mois de la classification statistique socio-économique nationale (« National Statistics Socio-economic Classification » ou NS-SEC) et les modifications apportées à un certain nombre de questions classificatoires des enquêtes gouvernementales en Grande-Bretagne.

La figure 2 résume la conception de l'ensemble de données de l'étude CNR de 2001 :

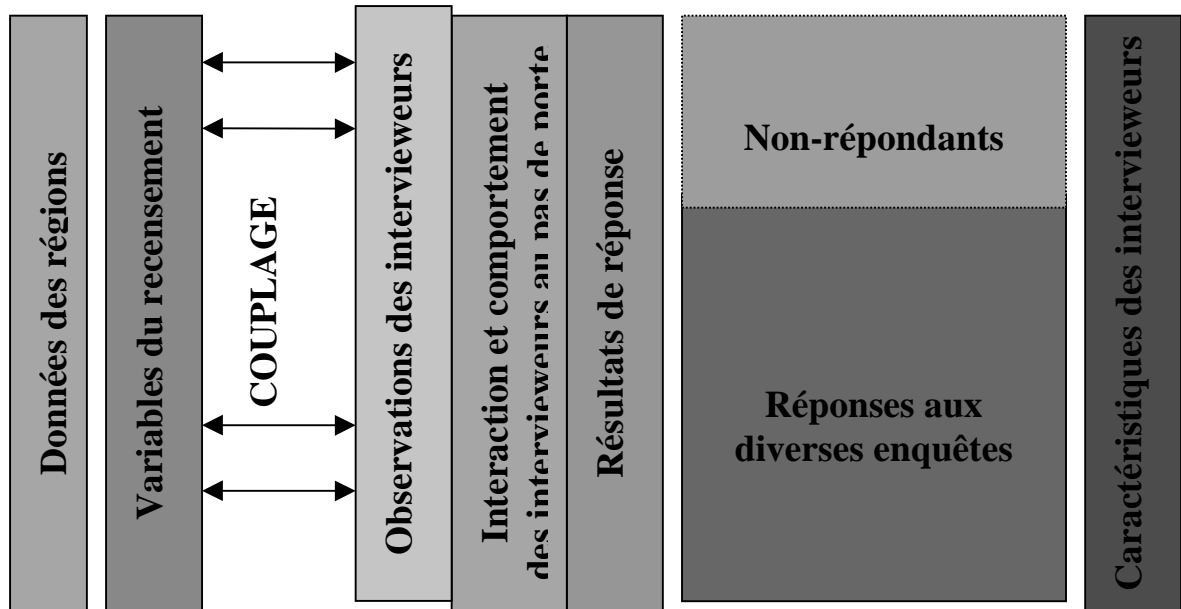


Figure 2 Conception de l'ensemble de données de l'étude CNR de 2001

5. STRATÉGIE D'ANALYSE

L'analyse portera sur deux aspects de la non-réponse, c'est-à-dire qu'on cherchera à savoir si l'intervieweur a pris contact avec l'enquêté ou non et si le ménage contacté a collaboré ou non. On tiendra compte de la structure hiérarchisée de l'ensemble de données dans l'analyse et, par une modélisation multiniveaux, on distinguera les effets propres aux niveaux considérés. Il y aura en gros trois niveaux d'information (figure 3) :

- intervieweurs – caractéristiques et attitudes;
- tâches – caractéristiques du plan d'enquête et caractéristiques des régions;
- ménages – observations et comportement des intervieweurs et caractéristiques des ménages.

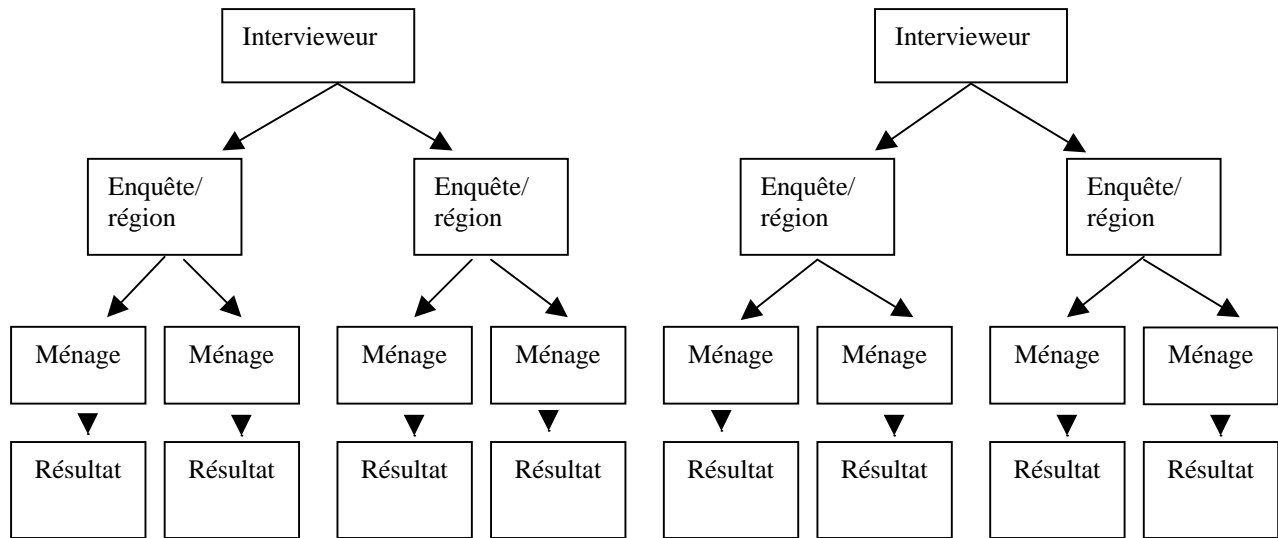


Figure 3 Structure des données

Les organismes de parrainage des enquêtes visées par l'étude paieront leur participation et commanderont des produits pour leurs enquêtes respectives. Comme dans l'étude de 1991, ils se verront offrir un éventail d'analyses à choisir. Outre des analyses par enquête, nous prévoyons un examen des incidences sur la non-réponse pour l'ensemble des enquêtes grâce à ce riche ensemble de données. Nous suivons en cela la stratégie appliquée pour une analyse récente des effets des caractéristiques des intervieweurs sur la non-réponse (Martin et Beerten, 1999).

Nous présenterons des rapports faisant état des résultats relatifs à une enquête à l'organisme qui parraine celle-ci. Il y aura aussi des rapports et des études méthodologiques où nous comparerons les résultats entre enquêtes et tirerons des conclusions plus générales sur la nature de la non-réponse.

Voici des possibilités d'analyse que nous offrirons :

- comparaisons descriptives des caractéristiques des ménages répondants, non joints et non répondants (refus de répondre);
- mesure du biais de non-réponse avec récapitulatifs des taux de surreprésentation ou de sous-représentation de sous-groupes pour les caractéristiques du recensement;
- comparaison des résultats précités avec les résultats des années antérieures (s'ils sont disponibles) afin d'identifier des variations dans le temps;
- modélisation de régression logistique afin de déterminer l'incidence relative de différentes variables du recensement sur les résultats de réponse;
- modélisation multiniveaux afin de déterminer les effets relatifs des variables de région, de ménage et (si elles sont disponibles ou appropriées) de conception d'enquête et d'intervieweur sur les résultats de réponse;
- plans de pondération élaborés en fonction des analyses qui précèdent (idéalement, nous emploierons des renseignements que peuvent couramment recueillir les intervieweurs, ainsi que des données servant aux corrections actuelles de non-réponse).

6. PROGRÈS DU PROJET À CE JOUR

Les travaux d'élaboration de cette étude ont débuté à la fin de 2000. Nous avons longuement conçu les procédures de collecte de données sur les observations des intervieweurs (caractéristiques ou état du secteur où se trouve une adresse, obstacles à l'accès aux ménages sélectionnés, interaction de l'intervieweur et de l'enquêté, par exemple). Un projet pilote a porté sur ces procédures en janvier et février 2001 avec les objectifs suivants :

- vérifier le bon fonctionnement des procédures de collecte de données dans une application simultanée à un certain nombre de grandes enquêtes permanentes;
- évaluer le temps que prennent les intervieweurs à recueillir les renseignements afin de s'assurer que le module de collecte de données peut s'insérer sans mal dans les tâches des intéressés;
- obtenir des suggestions d'amélioration de la conception en vue de réduire le plus possible le fardeau des interviews et de mieux concevoir le matériel pédagogique destiné à l'étape principale.

Ont participé à ce projet pilote dix intervieweurs travaillant à diverses enquêtes gouvernementales réalisées par l'ONS auprès des ménages. Ils ont donné de la rétroaction et des conseils au sujet de la formulation et de la disposition des questions, ainsi que sur des questions pratiques comme celle de l'utilisation du matériel sur le terrain. Voici les principales constatations faites dans le cadre du projet pilote :

- les intervieweurs pouvaient tirer les renseignements exigés de leurs observations ou de leur entretien initial avec l'enquêté;
- il leur fallait 10 minutes en moyenne pour consigner leurs observations à chaque adresse, et ce travail pouvait s'insérer dans leurs tâches;
- les renseignements nécessaires devaient être consignés le plus tôt possible, car ils pouvaient aisément s'oublier; les intervieweurs qui ont participé au projet pilote ont recommandé que cette consignation se fasse immédiatement après la prise de contact ou du moins le même jour;
- il fallait disposer d'une formule compacte et facile d'utilisation pour la consignation des observations des intervieweurs, car il était parfois dangereux ou peu commode d'entrer ces renseignements directement dans un ordinateur portatif lorsqu'on se retrouvait dans sa voiture après une visite.

Nous avons fait entrer ces constatations dans l'exercice de révision des procédures et communiqué la formule corrigée de consignation d'observations à trois des intervieweurs participants et à douze autres. Cette seconde consultation a permis d'intégrer certaines suggestions de simplification à la disposition de la formule. On a pu simplifier en conséquence la version définitive de cette formule.

La collecte des données sur les observations des intervieweurs a commencé en mai 2001. La collecte est terminée pour toutes les enquêtes visées sauf une. L'étude des intervieweurs a pris fin en août, et on a obtenu un taux de réponse de 84 % à cette enquête. Les données de recensement seront extraites en 2002 à des fins de couplage. Nous prévoyons entreprendre l'analyse en 2003.

BIBLIOGRAPHIE

- Barton, J. (1999), « Effective Calling Strategies for Interviewers on Household Surveys », *Survey Methodology Bulletin*, 44, pp. 14-26.
- Campanelli, P., Sturgis, P. et Purdon, S. (1997), *Can You Hear Me Knocking? An investigation into the impact of interviewers on survey response rates*, London: Social and Community Planning Research.
- Foster, K. (1998), *Evaluating Non-response on Household Surveys*, GSS Methodology Series No. 8, London: Government Statistical Service.
- Groves, R.M. and Couper, M. P. (1998), *Nonresponse in Household Interview Surveys*, New York: Wiley.
- Hox, J. J. (1999), « The Influence of Interviewer Attitudes and Behaviour on Household Survey Nonresponse: an International Comparison », article présenté à l'International Conference on Survey Nonresponse, Portland, Oregon.
- Lynn, P., Clarke, P., Martin, J. et Sturgis, P. (1999), « The Effects of Extended Interviewer Efforts on Nonresponse Bias », article présenté à l'International Conference on Survey Nonresponse, Portland, Oregon.

Martin, J. et Beerten, R. (1999), « The Effect of Interviewer Characteristics on Survey Response Rates », article présenté à l'International Conference on Survey Nonresponse, Portland, Oregon.

Morton-Williams, J. (1993), *Interviewer Approaches*, Cambridge: Dartmouth.