Recueil du Symposium 2014 de Statistique Canada
Au-delà des méthodes traditionnelles d'enquêtes : l'adaptation à un monde en évolution

Propriétés des mesures de la dépense énergétique quotidienne habituelle

Wayne A. Fuller[1] et Dave Osthus

Résumé

L'utilisation de moniteurs et l'autodéclaration sont deux méthodes de mesure de l'énergie dépensée durant l'activité physique, la variance de l'erreur étant habituellement beaucoup plus faible dans le cas des moniteurs que dans celui de l'autodéclaration. La Physical Activity Measurement Survey a été conçue pour comparer les deux procédures en utilisant des observations répétées sur une même personne. Ces observations répétées permettent de calibrer la mesure par autodéclaration sur la mesure par moniteur, ce qui rend possible l'estimation des composantes des variances des erreurs de mesure. Les estimations des composantes de la variance de l'erreur de mesure de la dépense d'énergie selon le moniteur et selon l'autodéclaration sont présentées pour les femmes qui ont participé à la Physical Activity Measurement Survey.

Mots-clés : Erreur de mesure, calibration, dépense énergétique habituelle, autodéclaration.

1. Introduction

L'activité physique est une composante importante de la dépense énergétique quotidienne moyenne de long terme, appelée dépense énergétique habituelle (DE habituelle). Les instruments d'autodéclaration et les appareils de monitorage, ou moniteurs, sont deux types d'instruments qui servent à mesurer la dépense énergétique quotidienne auprès de grands échantillons. L'utilisation d'un rapport personnel d'activité (RPA) est pratique dans les enquêtes à grande échelle, tandis que la plupart des procédures de monitorage sont trop coûteuses ou (et) trop astreignantes pour être employées pour tous les répondants à une grande enquête. Une procédure pratique consiste à utiliser le RPA pour tous les répondants et à sélectionner un sous-échantillon (ou un échantillon distinct) pour lequel on se sert du RPA ainsi que du moniteur pour recueillir les données auprès du même répondant pendant une même période. Les données répétées peuvent alors être utilisées pour construire une fonction de calibration reliant le RPA au moniteur.

Tant les moniteurs que les RPA donnent lieu à des erreurs de mesure. Dans le cas des moniteurs, les erreurs de mesure résultent de l'incapacité de l'appareil à saisir exactement la gamme complète d'activités et de la conversion imparfaite des données du moniteur en estimations de la dépense énergétique. Voir Welk (2002), ainsi que Welk et coll. (2004). Les erreurs de mesure associées aux instruments d'autodéclaration sont dues à des facteurs tels que les effets de désirabilité sociale, la difficulté à comprendre les concepts des questions de l'enquête, et la limitation cognitive de la remémoration d'activités qui ont eu lieu dans le passé. Voir Adams et coll. (2005), Troiano et coll. (2008), et Mathews (2002).

Nous présentons des procédures d'estimation qui conviennent pour des études pour lesquelles on dispose de plus d'une observation par méthode de mesure par personne. Nous calibrons le rapport personnel d'activité physique sur une mesure effectuée par moniteur et estimons les propriétés des erreurs de mesure.

[1] Wayne A. Fuller, Department of Statistics, Iowa State University, Ames, IA, 50011, waf@iastate.edu.
Dave Osthus, Department of Statistics, Iowa State University, Ames, IA, 50011.

2. La Physical Activity Measurement Survey (PAMS) de l'Iowa

La PAMS a été réalisée en continu pendant deux années dans quatre comtés de l'Iowa (Black Hawk, Dallas, Marshall et Polk) à partir de l'automne 2009. L'enquête a été réalisée auprès d'un échantillon probabiliste stratifié à plusieurs degrés comprenant deux strates par comté. Pour chaque comté, on a formé une strate à « forte prévalence de minorités » définie par les secteurs de recensement ayant un pourcentage relativement élevé de minorités et une strate à « faible prévalence de minorités » définie par les secteurs de recensement ayant un pourcentage relativement faible de minorités. Les strates à « forte prévalence de minorités » ont été suréchantillonnées afin d'obtenir un pourcentage plus élevé de minorités dans l'échantillon. Dans chaque strate, les ménages ont été sélectionnés systématiquement à partir de la liste de numéros de téléphone des pages blanches de l'annuaire. Une interview de présélection a été menée pour sélectionner aléatoirement dans chaque ménage un adulte satisfaisant aux critères de participation à l'enquête. Pour être admissible, l'adulte devait être âgé de 21 à 70 ans, capable de s'adonner à une activité physique et avoir la compétence requise pour répondre à l'interview. Après qu'il ait accepté de participer à l'étude, on a déterminé son poids et sa taille. Chaque répondant a fourni des données pendant deux jours. Les jours de mesure ont été attribués aléatoirement, à environ deux semaines d'intervalle. Durant les jours de mesure qui lui étaient attribués, le répondant a porté le moniteur d'activité (brassard SenseWear) pendant 24 heures, sauf au cours d'activités aquatiques, comme la natation ou la douche. Pendant les périodes où il ne portait pas le brassard, le répondant a inscrit ses activités dans un journal. Le lendemain du jour de mesure, le répondant a été contacté par téléphone et a produit un RPA portant sur la même période de 24 heures. L'instrument d'interview, conçu pour une enquête par sondage téléphonique, était fondé sur Mathews et coll. (2002).

Notre analyse porte sur les 785 femmes comprises dans l'échantillon de la PAMS qui ont fourni les deux mesures de dépense énergétique pour les deux jours de mesure. Plus de 90 % des répondantes (735 sur 785) dans l'échantillon ont porté le moniteur pendant plus de 90 % de la journée. L'activité durant les périodes où le moniteur n'avait pas été porté a été estimée en se basant sur le journal tenu par chaque répondante. Une observation pour laquelle les écarts étaient extrêmes a été exclue des analyses.

Les mesures de la dépense énergétique (DE) quotidienne en kilocalories par jour (kcal/j) ont été fournies par le moniteur. Les activités déclarées dans le RPA ont été classées selon une liste de 270 activités. À chacune de ces activités est attribué un niveau d'intensité exprimé en équivalent métabolique (CMET) en utilisant une version modifiée du Compendium of Physical Activities. Voir Ainsworth et coll. (1993) et Ainsworth et coll. (2000). Une version du compendium est disponible sur le Web sous l'entrée « The Compendium of Physical Activities Tracking Guide ». Un CMET est l'énergie que requiert un « individu standard » pour une activité par rapport à l'activité de référence consistant à être assis calmement. Un MET est une mesure de la dépense énergétique par rapport au poids, où 1 MET = 0,0175 kcal/kg/min. Le nombre total de CMET pendant un jour est égal à la somme des produits $\text{CMET}_a \times \text{durée}_a$, où durée_a est le temps consacré à l'activité a, et la somme est effectuée sur l'ensemble des activités.

3. Modèle

Les données de base sont la DE selon le moniteur pour les 24 heures du jour de mesure et le nombre total de CMET pour le même jour. Chaque personne fournit deux paires de données. La calibration du RPA aux données du moniteur requiert que les deux mesures soient exprimées dans la même unité. La DE, qui est l'unité d'intérêt en recherche sur l'activité physique, est obtenue directement du moniteur. Par conséquent, le nombre observé de CMET calculé d'après les réponses personnelles doit être transformé en DE. Une procédure consiste à multiplier le CMET par une fonction du poids, de la taille et de l'âge, la plus fréquemment utilisée étant l'équation de Harris-Benedict. L'équation originale de Harris-Benedict pour les femmes est

$$\text{DER} = 655 + (9{,}6 \times \text{ poids en kg}) + (1{,}8 \times \text{ taille en cm}) - (4{,}7 \times \text{ âge en années}),$$

où DER est la dépense énergétique au repos, c'est-à-dire les calories brûlées par une personne non active pendant une période de 24 heures. Comme il existe de nombreuses autres formules et que la formule de Harris-Benedict peut ne pas être applicable directement à la DE mesurée par le moniteur, nous estimons une équation analogue à l'équation de Harris-Benedict à travers l'estimation par calibration.

Les caractéristiques personnelles ont une incidence sur l'activité, la dépense énergétique et l'autodéclaration. Dans certaines études, les personnes en surpoids ont eu tendance à surdéclarer leur activité physique comparativement aux personnes de poids normal. En outre, les données semblent indiquer que certains de ces facteurs ont également une incidence sur les données du moniteur. Pour élaborer une équation de calibration, la façon dont les caractéristiques personnelles influent sur l'autodéclaration présente un intérêt, mais nous pouvons seulement estimer cet effet par rapport à un autre instrument. Par conséquent, le coefficient estimé d'une caractéristique personnelle dans une fonction reliant les CMET du RPA à la DE du moniteur peut être composé de trois effets, à savoir l'effet sur la déclaration, l'effet sur le moniteur et l'effet de la conversion de l'activité en DE. Néanmoins, nous considérons notre fonction estimée comme un estimateur de la DER.

Soit C_{ij} la valeur totale des CMET pour l'individu *i* le jour *j* et soit

$$DE_{C,ij} = (\sum_{h=1}^{4} \alpha_h z_{hi}) C_{ij}, \qquad (1)$$

où $\text{z}_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i}) = (1, \text{poids en kg, taille en cm, âge en années})$ pour l'individu *i* et le vecteur $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$ doit être estimé.

Les variances des erreurs de mesure pour les deux procédures sont corrélées à la valeur et les distributions sont asymétriques. Les données logarithmiques sont nettement moins asymétriques que les données originales et les variances paraissent moins corrélées à la valeur. Pour des exemples de l'utilisation des logarithmes en recherche sur l'activité physique, voir Ferrai et coll. (2007) et Tooze et coll. (2013). Dans notre modèle, nous employons les logarithmes, mais d'une manière un peu différente de celle de Tooze et coll. (2013).

Supposons que les observations satisfont le modèle

$$\begin{aligned} DE_{M,ij} + \zeta_M &= k_{ij}^* u_{ij}^*, \\ DE_{C,ij} + \zeta_C &= \beta_0^* k_{ij}^* r_i^* e_{ij}^*, \end{aligned} \qquad (2)$$

où $k_{ij}^* - \zeta_M$ est la dépense énergétique « réelle » inobservable de l'individu *i* le jour *j*, $(\zeta_M, \zeta_C, \beta_0^*)$ est un vecteur de paramètres et $(u_{ij}^*, r_i^*, e_{ij}^*)$ est un vecteur de variables aléatoires. Nous représentons l'exponentiation par * de sorte que, par exemple, $u_{ij} = \log u_{ij}^*$. Le modèle d'erreur de mesure proposé est

$$\begin{aligned} x_{ij}(\zeta_M) &= \log(DE_{M,ij} + \zeta_M) = \mu_x + t_i + d_{ij} + u_{ij}, \\ y_{ij}(\zeta_C) &= \log(DE_{C,ij} + \zeta_C) = \mu_y + t_i + d_{ij} + r_i + e_{ij}, \end{aligned} \qquad (3)$$

où

- μ_x est la moyenne de population de *x*;
- μ_y est la moyenne de population de *y*;
- $t_i \sim (0, \sigma_t^2)$ est l'écart de la moyenne de l'individu *i* par rapport à la moyenne de population de *x*;
- $d_{ij} \sim (0, \sigma_d^2)$ est l'écart du jour *j* par rapport à la moyenne de l'individu *i*;
- $u_{ij} \sim (0, \sigma_u^2)$ est l'erreur de mesure du moniteur;
- $r_i \sim (0, \sigma_r^2)$ est l'erreur moyenne de déclaration pour l'individu *i*;
- $e_{ij} \sim (0, \sigma_e^2)$ est l'erreur de mesure individuelle dans l'instrument d'autodéclaration.

Le terme $k_{ij} = \mu_x + t_i + d_{ij}$ est la valeur « réelle » inobservable de *x* sur l'échelle transformée. Nous supposons que les variables aléatoires $(t_i, d_{ij}, u_{ij}, e_{ij}, r_i)$ sont mutuellement indépendantes. Notons que la transformation inverse d'une valeur prédite par le modèle (3) résulte en une fonction linéaire de la valeur originale.

4. Estimation

Nous avons estimé le vecteur $\boldsymbol{\alpha}$ de (1) par régression de $\mathrm{C}_{ij}^{-1} DE_{M,ij}$ sur $\mathbf{z}_i$ à l'aide de la méthode des moindres carrés généralisés (MCG). Les coefficients estimés donnés au tableau 1 sont similaires à ceux de l'équation de Harris-Benedict, mais diffèrent de manière significative. Les écarts entre les estimations et les coefficients de Harris-Benedict sont donnés dans la troisième colonne du tableau 1. Les erreurs types sont celles estimées par les MCG. La statistique F du test de l'hypothèse que $(\alpha_2, \alpha_3, \alpha_4)$ est égal au vecteur de Harris-Benedict vaut 8,52 avec 3 et 781 degrés de liberté.

Tableau 1
Coefficients estimés de l'équation de la DER

Terme	Estimation	Estimation - $\alpha(HB)$	Erreur type
α_1 (ordonnée à l'origine)	-78,31	-733,32	177,78
α_2 (poids)	8,70	-0,91	0,35
α_3 (taille)	6,29	4,49	1,07
α_4 (âge)	-3,18	1,51	0,62

Nous avons utilisé le vecteur $\boldsymbol{\alpha}$ estimé pour construire $DE_{C,ij}$ de (1). Nous avons fixé ζ_C à -350, soit une valeur telle que la variance de $\log(DE_{C,ij})$ est presque une fonction constante de $\log(DE_{C,ij})$. L'estimation de ζ_M est la valeur de ζ_M telle que le β estimé pour un modèle étendu est égal à un, où le modèle étendu est

$$\begin{aligned} x_{ij}(\zeta_M) &= \mu_x + t_i + d_{ij} + u_{ij}, \\ y_{ij}(\zeta_C) &= \mu_y + \beta(t_i + d_{ij}) + r_{ij} + e_{ij}. \end{aligned} \tag{4}$$

Sachant ζ_C et ζ_M, nous avons appliqué la méthode des moments pour estimer le vecteur de paramètres $(\beta, \sigma_t^2, \sigma_d^2, \sigma_u^2, \sigma_e^2, \sigma_r^2)'$. Nous avons construit les moments pour le vecteur des observations individuelles,

$$\mathrm{a}_i = (\bar{x}_i, \bar{y}_i, x_{i1} - x_{i2}, y_{i1} - y_{i2})', \tag{5}$$

où $(\bar{x}_i, \bar{y}_i) = [0,5(x_{i1} + x_{i2}), 0,5(y_{i1} + y_{i2})]$, $x_{ij} = \log(DE_{M,ij} + \hat{\zeta}_M)$, $y_{ij} = \log(\hat{D}E_{C,ij} + \zeta_C)$, et $\hat{D}E_{C,ij}$ est définie par (1) en remplaçant $\boldsymbol{\alpha}$ par $\hat{\boldsymbol{\alpha}}$. Étant donné les hypothèses du modèle et les paramètres connus, $E\{\mathrm{a}_i\} = 0$ et la variance de $\mathbf{a}_i$ est

$$V\{\mathrm{a}_i\} = \begin{pmatrix} \sigma_t^2 + 2^{-1}(\sigma_d^2 + \sigma_u^2) & \beta\sigma_t^2 + 2^{-1}\beta\sigma_d^2 & 0 & 0 \\ & \beta^2(\sigma_t^2 + 2^{-1}\sigma_d^2) + \sigma_r^2 + 2^{-1}\sigma_e^2 & 0 & 0 \\ & & 2(\sigma_d^2 + \sigma_u^2) & 2\beta\sigma_d^2 \\ \textit{symétrique} & & & 2(\beta^2\sigma_d^2 + \sigma_e^2) \end{pmatrix}. \tag{6}$$

La variance d’échantillon simple de $\mathbf{a}_i$ est

$$\hat{V}\{\mathrm{a}_i\} = (n-1)^{-1} \sum_{i=1}^{n} (\mathrm{a}_i - \bar{\mathrm{a}})(\mathrm{a}_i - \bar{\mathrm{a}})', \tag{7}$$

où $\bar{\mathrm{a}} = n^{-1} \sum_{i=1}^{n} \mathrm{a}_i$ est la moyenne d’échantillon de $\mathbf{a}_i$ et n est la taille d’échantillon. Les estimateurs par la méthode des moments sont donnés en posant $\hat{V}\{\mathrm{a}\} = V\{\mathrm{a}\}$ et en solutionnant pour trouver les paramètres de (6). Ces équations contiennent six paramètres du modèle et six moments uniques, ce qui permet d’exprimer chaque paramètre estimé du modèle sous forme d’une fonction des moments de l’échantillon.

Les estimations des paramètres du modèle sont données au tableau 2. Les variances des estimations du tableau 2 ont été calculées en utilisant une procédure jackknife avec suppression d’un groupe, telle qu’elle est décrite dans Kott (2001) et à la section 4.2.2 de Fuller (2009). Nous avons formé 25 partitions en classant d’abord les individus en fonction de leur race et de la strate d’échantillonnage. Dans chaque combinaison race/strate, nous avons classé les individus par âge et nous avons attribué systématiquement les numéros de partition 1 à 25 aux individus classés par ordre. Nous avons créé une réplique en supprimant une des partitions. La variance estimée par le jackknife d’un estimateur $\hat{\theta}$ est

$$\hat{V}\{\hat{\theta}\} = \frac{1}{B(B-1)} \sum_{b=1}^{B} (\hat{\theta}_b - \hat{\theta}_{JK})(\hat{\theta}_b - \hat{\theta}_{JK})'$$

où $\hat{\theta}_{JK} = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b$, $\hat{\theta}_b = B\hat{\theta} - (B-1)\hat{\theta}^{(b)}$, $\hat{\theta}^{(b)}$ est l’estimation calculée pour la b^{e} réplique, et B = 25. Tous les calculs, y compris l’estimation de $\boldsymbol{\alpha}$, ont été effectués pour chaque réplique. Comme dans l’estimation originale, nous avons traité ζ_C comme étant fixe.

D’après les estimations du tableau 2, environ 15 % de la variation de la DE_M observée sont dus à l’erreur de mesure et environ 17 %, à la variation d’un jour à l’autre de la dépense énergétique individuelle. La variation d’une personne à l’autre de la *DE* habituelle représente environ 68 % de la variance totale. Le σ_u de 0,664 est approximativement équivalent à un écart-type de 179 sur l’échelle originale.

L’erreur de mesure dans le RPA est égale à la somme des variances de *r* et de *e* du modèle (3). La somme des deux variances est égale à environ 95 % de la variance de la *DE* quotidienne (environ 1,18 % de la *DE* habituelle). Donc, une détermination individuelle par moniteur possède un écart-type relatif d’environ 32 % en tant qu’estimateur de la *DE* habituelle, et la *DE* selon le RPA calibré possède un écart-type d’environ 1,18 % en tant qu’estimateur de la *DE* habituelle.

L’interprétation des variances du tableau 2 doit se faire en tenant compte du fait que $DE_{C,ij}$ et $DE_{M,ij}$ sont toutes deux des fonctions du poids, de la taille et de l’âge. Si les erreurs dans $DE_{M,ij}$ sont corrélées à ces facteurs, la variance estimée de l’erreur pour DE_M et la covariance estimée entre DE_C et DE_M sont affectées.

L’estimation de $\exp(\mu_y - \mu_x)$ est une estimation de β_0^* du modèle (2). Ce paramètre est la pente sur l’échelle originale de la régression de DE_C sur la *DE* réelle, où cette dernière est la valeur probable de la mesure du moniteur.

Tableau 2
Paramètres estimés du modèle

Paramètre	Estimation	Erreur type
ζ_M	346	150
$\beta_0 = \mu_y - \mu_x$	-0,299	0,017
$\exp(\mu_y - \mu_x)$	0,742	0,043
$100\sigma_t^2$	1,975	0,255
$100\sigma_d^2$	0,478	0,063
$100\sigma_u^2$	0,441	0,135
$100\sigma_e^2$	1,130	0,123
$100\sigma_r^2$	1,200	0,093

Le passage de l'échelle logarithmique à l'échelle originale doit se faire avec soin, parce que le logarithme de l'espérance d'une variable aléatoire n'est pas égal à l'espérance des logarithmes. Notre représentation rend la transformation plus facile que dans la plupart des cas, parce que les éléments de (2) et (3) sont des fonctions relativement simples. Supposons que l'on souhaite effectuer la calibration sur l'échelle originale et que l'on désire que l'espérance conditionnelle (à la vraie valeur) du RPA calibré soit égale à la valeur probable conditionnelle du moniteur. En utilisant (2), la DE selon le RPA calibré est

$$DE_{RPA} = (\beta_0^*)^{-1}\zeta_C - \zeta_M + (\beta_0^*)^{-1} DE_{C,ij}\,. \tag{8}$$

Les estimations provenant du tableau 2 donnent

$$DE_{RPA} = -817,7 + 1,348\, DE_{C,ij}. \tag{9}$$

Un autre estimateur de $(\beta_0^*)^{-1}$ est

$$(\tilde{\beta}_0^*)^{-1} = \left[\sum_{i=1}^{n}\sum_{j=1}^{2} (DE_{C,ij} + \hat{\zeta}_C)\right]^{-1} \sum_{i=1}^{n}\sum_{j=1}^{2} (DE_{M,ij} + \zeta_M) = 1,338. \tag{10}$$

Le deuxième estimateur, qui est calculé sur l'échelle originale, devrait être moins biaisé que l'estimateur transformé à partir de l'échelle logarithmique. Dans le cas présent, les deux estimations sont très similaires.

Bibliographie

Adams, S. A., C. E. Matthews, C. B. Eberling, C. G. Moore, J. E. Cunningham, J. Fulton et J. R. Hebert (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology*; **161**: 389-398.

Ainsworth, B. E., W. L. Haskel, M. C. Whitt, M. L. Irwin, A. M. Swartz, S. J. Strath, W. L. O'Brien, D. R. Bassett, K. H. Schmitz, P. O. Emplaincourt, D. R. Jacobs, A. S. Leon (2000). Compendium of physical activities: An update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*; **32**: S498-S516.

Ainsworth B. E., W. L. Haskell, A. S. Leon, D. R. Jacobs, H. J. Montoye, J. F. Sallis, R. S. Paffenbarger (1993). Compendium of physical activities: Classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise*; **25**: 71-80.

Ferrari, P., C. Friedenreich, C. E. Matthews (2007). The role of measurement error in estimating levels of physical activity. *American Journal of Epidemiology*; **166**: 832-840.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley: New York, 20-25.

Fuller, W. A. (2009). *Sampling Statistics*. Wiley: New York; 251-260.

Kott, P. S. (2001). The delete-one-group jackknife. *Journal of Official Statistics*: **17**: 521-526.

Matthews, C. E. (2002). Use of self-report instruments to assess physical activity. Dans Welk, G. J. *Physical Activity Assessments for Health Related Research.* Human Kinetics: Champaign, 107-123.

Tooze, J., R. Troiano, R. Carroll, A. Moshfegh, L. Freedman (2013). A measurement error model for physical activity level as measured by a questionnaire with application to the 1999-2006 NHANES questionnaire. *American Journal of Epidemiology;* **177**: 1199-1208.

Troiano, R. P., D. Berrigan, K. W. Dodd, L. C. Masse, T. Tilert, M. McDowell (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise;* **40**: 181-188.

Van Remoortel, H. et coll. (2012). Validity of activity monitors in health and chronic disease: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*; **9**: 84.

Welk, G. J. (2002). Introduction to physical activity research. Dans Welk G. J. *Physical Activity Assessments for Health Related Research.* Human Kinetics: Champaign, 3-18.

Welk, G. J., J. A. Schaben, J. R. Morrow (2004). Reliability of accelerometry-based activity monitors: A generalizability study. *Medicine and Science in Sports and Exercise;* **36**: 1637-1645.