

Traitement des données de consommations alimentaires
Données mises en forme dans le cadre du projet ANR-12-DSSA-0001 SOFI
B de Lauzon-Guillain, INSERM CRESS équipe ORCHAD
blandine.delauzon@inserm.fr

INTRODUCTION

L'objectif de ce traitement des données était d'identifier les sujets qui avaient un questionnaire de fréquences de consommations alimentaires exploitable, et pour ces sujets, de calculer un apport journalier en grammes/ml pour chacun des items du questionnaire ainsi qu'un apport nutritionnel journalier (énergie et nutriments).

GESTION DES DONNEES MANQUANTES

Pour les fréquences de consommation des aliments et boissons non alcoolisées

- 11661 sans données manquantes (DM) sur les 122 aliments nécessaires pour le calcul de l'apport nutritionnel
- 2146 sujets n'ont pas complété le FFQ
- Nb données manquantes sur les fréquences
 - a. 2142 avec 1 DM
 - b. 673 avec 2 DM
 - c. 1024 avec 3-10 DM
 - d. 667 avec >10 DM

Un questionnaire a été considéré comme non exploitable lorsqu'il avait plus de 10 valeurs manquantes. L'apport énergétique n'a donc pas été calculé pour les 667 sujets concernés. Pour les sujets ayant au plus 10 valeurs manquantes, les valeurs manquantes ont été remplacées par la médiane de fréquence de consommation des aliments concernés.

Pour les portions

- 14598 sujets sans portion manquante pour les aliments
- 2340 sujets avec aucune portion pour les aliments (2146 n'ont pas rempli le FFQ)
- 937 avec 1 portion manquante, 438 avec 2-13 portions manquantes pour les aliments
- La majorité des sujets ont de nombreuses données manquantes pour les boissons

Les données manquantes sur les tailles des portions ont été remplacées par la portion médiane.

CALCUL DES FREQUENCES JOURNALIERES

Pour permettre un calcul l'apport journalier de chaque aliment ou groupe d'aliment du questionnaire, les catégories de fréquences ont été transformées en **fréquence moyenne annuelle** de la façon suivante :

Pour la majorité des aliments

Les fréquences déclarées ont été transformées de la manière suivante :

- Jamais → 0/an
- Moins d'une fois par mois → 4 fois/an
- Entre 1 et 3 fois par mois → 24 fois / an (valeur médiane de l'intervalle)
- 1 fois par semaine → 52 fois/an
- Entre 2 et 5 fois par semaine → 182 fois par an (valeur médiane de l'intervalle)
- 1 fois par jour ou presque → 365,25 fois/an
- Plusieurs fois par jour → 730,5 fois par an

Pour les familles d'aliments avec beaucoup d'items

Les résultats de l'étude de validation (soumise pour publication) ont amené à faire des modifications sur la transformation de certaines fréquences pour limiter la surestimation liée au nombre d'items de certaines familles d'aliments. Ainsi, il a été décidé pour les familles d'aliments avec plus de 8 items que les fréquences seraient transformées de la manière suivante :

- Jamais → 0/an
- Moins d'une fois par mois → 4 fois/an
- Entre 1 et 3 fois par mois → **12** fois / an (valeur inférieure de l'intervalle)
- 1 fois par semaine → 52 fois/an
- Entre 2 et 5 fois par semaine → 104 fois par an (valeur inférieure de l'intervalle)
- 1 fois par jour ou presque → 365,25 fois/an
- Plusieurs fois par jour → 730,5 fois par an

Cela concerne les items des familles d'aliments suivantes :

- Légumes crus + Légumes cuits (17 items au total)
- Féculents et céréales (10 items)
- Viande/Jambon/Œuf (9 items)
- Poisson (9 items)
- Plats cuisinés (12 items)
- Fruits (10 items)

Pour les boissons non alcoolisées

Les fréquences déclarées ont été transformées de la manière suivante :

- Jamais → 0/an
- Moins d'un verre par mois → 4 verres/an
- Entre 1 et 3 verres par mois → 24 verres / an (valeur médiane de l'intervalle)
- 1 verre par semaine → 52 verres/an
- Entre 2 et 5 verres par semaine → 182 verres par an (valeur médiane de l'intervalle)
- 1 verre par jour ou presque → 365,25 verres/an
- 2 à 4 verres par jour → 1095,75 verres/an (valeur médiane de l'intervalle)
- 4 à 8 verres par jour → 2191,5 verres/an (valeur médiane de l'intervalle)
- Plus de 8 verres par jour → 3287,25 verres par an (ce qui correspond à 9 verres par jour)

CALCUL DE L'APPORT JOURNALIER POUR CHAQUE ITEM DU QUESTIONNAIRE

Table des portions

Les correspondances entre les photos incluses dans le questionnaire de fréquences de consommation s'appuient sur le portionnaire SUVIMAX (Hercberg et al., 2002). Un premier rapprochement entre les items du FFQ et la taille des portions en grammes a été réalisé, dans le cadre de l'étude de validation du FFQ. Il a été ensuite repris pour adapter ce regroupement aux quelques modifications apportées sur le questionnaire entre l'étude de validation et l'enquête nationale.

Des tailles de portions ont été recueillies pour 15 aliments et 12 boissons. Elles ont été utilisées pour 76 aliments et boissons. Par exemple, la photo « crudités en entrée » a été utilisée pour les 4 items « légumes crus » et la photo « Steak » a été utilisée pour les 5 items « viande hors volaille ». Pour les 50 items pour lesquels aucun rapprochement avec une photo n'a pu être fait, la portion moyenne du portionnaire SUVIMAX (portion D) a été attribuée.

Le détail de tous les regroupements est présenté dans un fichier Excel (Construction_Base_Portion).

Apport en grammes

L'apport en g (pour les aliments) ou en ml (pour les boissons) a été calculé pour chaque item du questionnaire en multipliant la fréquence journalière par la portion attribuée à l'item (selon la photo choisie ou la portion moyenne)

CALCUL DE L'APPORT EN NUTRIMENTS POUR CHAQUE INDIVIDU

Table de composition nutritionnelle

La table de composition alimentaire issue de l'étude SUVIMAX a été traitée et adaptée à la version « Enquête Nationale » du FFQ Elfe (Hercberg, 2006). La table SUVIMAX a été préférée à la table CIQUAL 2013 car elle permet de distinguer les protéines végétales des protéines animales, permet de calculer un apport pour les acides gras majeurs et comporte moins de données manquantes que la table CIQUAL 2013.

Principe de regroupement des aliments SUVIMAX :

- Sélection de tous les aliments SUVIMAX qui peuvent être inclus dans un item du FFQ
- Pour les items du FFQ qui ne sont pas mentionnés dans la table SUVIMAX (semoule, barres céréalières, desserts fermentés au soja du rayon frais, boisson au soja), les compositions des items correspondants dans la Table de composition nutritionnelle des aliments CIQUAL 2013 ont été pris en compte et ajoutées à ceux de la table SUVIMAX.
- Pour un item du questionnaire regroupant plusieurs aliments (exemple : « des biscottes, des cracottes, des pains grillés (type suédois,...) ») : calcul de la moyenne non pondérée sur tous les aliments pour estimer l'apport d'un item du FFQ pour un nutriment donné.

L'item « aliments allégés » est utile pour la détermination de profil de consommation mais pas pour l'estimation des apports journaliers. En effet, aucune composition nutritionnelle n'a été indiquée pour les aliments allégés, cet item n'existant dans aucune table de composition

nutritionnelle publiée et la nature de l'allègement n'étant pas précisée (en sel, en gras, en sucre...). Les aliments allégés sont donc pris en compte pour l'apport en gramme de chaque groupe d'aliment, mais pas pour le calcul de l'apport énergétique total. Cette décision a vraisemblablement peu d'impact sur l'apport énergétique total dans la mesure où 14200/15813 (89,8%) femmes n'en consomment jamais et 889 (5,6%) en consomment moins d'1 fois par mois.

Le détail de tous les regroupements est présenté dans un fichier Excel (BaseCompo_SUVIMAX_Mai2014).

Apport en nutriments

L'apport en un nutriment j (hors alcool) a été calculé de la manière suivante :

$$\sum_{i=1}^{i=122} (\text{Apport en aliment } i \text{ (en g)} * \text{Apport nutriment } j \text{ pour } 100\text{g d'aliment } i) / 100$$

Les apports journaliers ont été calculés (avec ou sans prise en compte des boissons alcoolisées) pour les nutriments suivants :

- Energie (kcal)/jour
- Energie (kJ)/jour
- Eau (g)/jour
- Proteines (g)/jour
- Proteines vegetales (g)/jour
- Proteines animales (g)/jour
- Lipides (g)/jour
- Acides Gras Satures (g)/jour
- Acides Gras MonoInsatures (g)/jour
- Acides Gras PolyInsatures (g)/jour
- 18:2 n-6 linoleique (g)/jour
- 18:3 n-3 alpha linolenique (g)/jour
- 20:4 n-6 arachidonique (g)/jour
- 20:5 n-3 EPA (g)/jour
- 22:5 n-3 DPA (g)/jour
- 22:6 n-3 DHA (g)/jour
- Cholesterol (mg)/jour
- Glucides totaux (g)/jour
- Glucides Simples (g)/jour
- Glucides simples ajoutés (g)/jour
- Glucides Complexes (g)/jour
- Fibres (g)/jour
- Fibres solubles (g)/jour
- Alcool (g)/jour
- Vit A - retinol (µg)/jour
- Beta-carotene (µg)/jour
- Vit B1 (mg)/jour
- Vit B2 (mg)/jour
- Vit B3 (mg)/jour
- Vit B5 (mg)/jour
- Vit B6 (mg)/jour
- Vit B9 (µg)/jour
- Vit B12 (µg)/jour
- Vit C (mg)/jour
- Vit D (µg)/jour
- Vit E (mg)/jour
- Calcium (mg)/jour
- Fer (mg)/jour
- Iode (µg)/jour
- Potassium (mg)/jour
- Magnesium (mg)/jour
- Sodium (mg)/jour
- Phosphore (mg)/jour

GESTION DES DONNEES ABERRANTES

L'application des formules d'identification des sur ou sous-déclarants (Goldberg et al., 1991, Black, 2000) est peu adaptée à l'utilisation de questionnaires de fréquences de consommations alimentaires. Par ailleurs, il n'existe pas de formule applicable chez les femmes enceintes.

Par conséquent, pour ne pas tenir compte des apports énergétiques improbables, il a été décidé d'exclure les femmes avec un apport énergétique hors alcool inférieur au 3^{ème} percentile (933 kcal) ou supérieur au 97^{ème} percentile (5072 kcal).

Par conséquent, le fichier contenant l'apport en grammes des différents groupes d'aliments, ainsi que l'apport en énergie et en nutriments ne comporte des valeurs que pour les femmes dont l'apport énergétique se situe entre le 3^{ème} et le 97^{ème} percentile, soit 14 556 femmes.

TRAITEMENT DE LA VARIABLE ALCOOL

Rappel du questionnaire

45. Boissons alcoolisées :

a. Combien avez-vous consommé de verres **PAR MOIS**, au cours de la période où vous étiez enceinte sans le savoir ?

- pas d'alcool durant cette période 0
- moins d'un verre par mois durant cette période 1
- au moins un verre par mois durant cette période 2

Si vous avez consommé plus d'un verre par mois, indiquez le nombre de verres au total **PAR SEMAINE**, week-ends compris. **Donnez un chiffre rond. Par exemple, pour moins d'un verre par semaine : indiquez 1.**

- de cidre ou de bière
- de vin blanc, rouge ou rosé
- d'apéritifs (pastis, cherry, porto, martini,...)
- d'alcools forts (whisky, gin, vodka, premix,...)

b. Combien avez-vous consommé de verres **PAR MOIS**, depuis que vous avez su que vous étiez enceinte ?

- pas d'alcool durant cette période 0
- moins d'un verre par mois durant cette période 1
- au moins un verre par mois durant cette période 2

Si vous avez consommé plus d'un verre par mois, indiquez le nombre de verres au total **PAR SEMAINE**, week-ends compris. **Donnez un chiffre rond. Par exemple, pour moins d'un verre par semaine : indiquez 1.**

- de cidre ou de bière
- de vin blanc, rouge ou rosé
- d'apéritifs (pastis, cherry, porto, martini,...)
- d'alcools forts (whisky, gin, vodka, premix,...)

Le remplissage se fait donc en 2 étapes, avec un filtre, la consommation détaillée par alcool n'est sensée être donnée qu'en cas de consommation au moins égale à 1 verre par mois. Néanmoins, le détail par type d'alcool est nécessaire pour le calcul nutritionnel.

Descriptif des réponses avant retraitement

Identification des incohérences/valeurs manquantes pour la variable début de grossesse :

- 2649 valeurs manquantes pour toutes les variables alcool
- 20 valeurs manquantes sur la consommation globale mais avec un détail sur au moins 1 des alcools (dont 14 avec le détail pour les 4 types d'alcool)
- Parmi les 2936 sujets qui ont complété le détail des alcools :
 - 2083 ont rempli les 4 questions
 - 97 en ont rempli 3
 - 260 en ont rempli 2
 - 496 en ont rempli 1 seul
- Parmi les 2916 sujets qui ont complété la question globale et le détail des alcools
 - 39 sujets ont indiqué ne pas avoir consommé, mais ont une consommation détaillée d'au moins 12 verres/an
 - 60 sujets ont indiqué avoir consommé moins d'un verre/mois, mais ont une consommation détaillée d'au moins 12 verres/an
 - 34 sujets ont indiqué avoir consommé au moins un verre/mois, mais ont une consommation détaillée < 12 verres/an

Identification des incohérences/valeurs manquantes pour la variable fin de grossesse :

- 3215 valeurs manquantes pour toutes les variables alcool
- 28 valeurs manquantes sur la consommation globale mais avec un détail sur au moins 1 des alcools (dont 22 avec le détail pour les 4 types d'alcool)
- Parmi les 564 sujets qui ont complété le détail des alcools :
 - 366 ont rempli les 4 questions
 - 11 en ont rempli 3
 - 52 en ont rempli 2
 - 135 en ont rempli 1 seul
- Parmi les 564 sujets qui ont complété la question globale et le détail des alcools
 - 33 sujets ont indiqué ne pas avoir consommé, mais ont une consommation détaillée d'au moins 12 verres/an
 - 12 sujets ont indiqué avoir consommé moins d'un verre/mois, mais ont une consommation détaillée d'au moins 12 verres/an
 - 8 sujets ont indiqué avoir consommé au moins un verre/mois, mais ont une consommation détaillée < 12 verres/an

Règles de décision pour gérer ces incohérences et avoir des données par type d'alcool

1. si incohérence, on privilégie la réponse au détail des alcools sur la réponse à la question générale (partant de l'hypothèse que les femmes ont tendance à sous-déclarer leur consommation)
2. Si au moins un type d'alcool est complété, les valeurs manquantes sur les autres types d'alcool sont mises à 0
3. Si le détail des alcools n'est pas complété,
 - a. Si la consommation générale est nulle -> tous les types d'alcool sont mis à 0
 - b. Si la consommation générale est <1/mois -> tous les types d'alcool sont mis à 1 verre/an
 - c. Si la consommation générale est ≥ 1 /mois -> tous les types d'alcool sont mis à 6 verres/an

Pour pouvoir calculer une consommation moyenne sur toute la grossesse, le 1^{er} trimestre est considéré comme une période « enceinte sans le savoir » et les 2 derniers trimestres comme « depuis que vous avez su que vous étiez enceinte ».

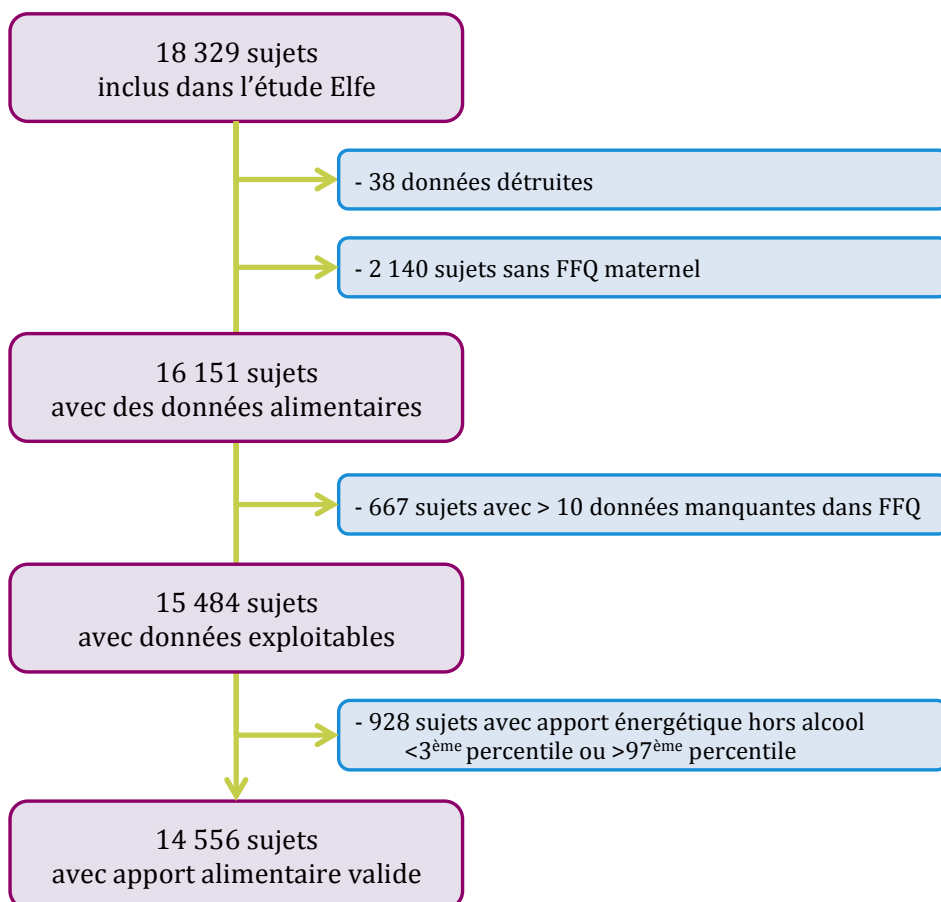
Calcul de la fréquence annuelle

Les fréquences déclarées pour chacune des périodes (avant de connaître la grossesse, en sachant la grossesse) ont été transformées de la manière suivante :

- Pas d'alcool pendant cette période → 0/an
- Moins d'un verre par mois durant cette période → 4 verres/an (1 pour chacun des types d'alcool proposés)
- Au moins 1 par mois
 - Si le nombre de verres par mois n'est pas précisé → 24 verres / an (6 pour chacun des types d'alcool proposés)
 - Si le nombre de verre par mois est précisé → chiffre indiqué

La fréquence moyenne pendant la grossesse a été calculée de la manière suivante : (fréquence au cours de la période où la femme était enceinte sans le savoir + 2*(fréquence au ou de la période où la femme savait qu'elle était enceinte))/3. Ce calcul a été fait sous l'hypothèse selon laquelle la période où la femme est enceinte sans le savoir correspond au 1^{er} trimestre de la grossesse.

SELECTION DE LA POPULATION



DESCRIPTIF DE LA BASE FINALE

La base finale comprend

- Variables indicatrices :
 - Indicateur selon le nombre de valeurs manquantes (ValMqt)
 - Apport énergétique <3^{ème} percentile ou >97^{ème} percentile (pbl_nrj)
 - questionnaire alimentaire valide (selectionFFQ)
- Chez les sujets avec un questionnaire alimentaire exploitable :
 - Fréquence journalière de consommation des 110 aliments, 12 boissons non alcoolisées et 4 boissons alcoolisées (alim1j → alim137j)
 - Quantités journalières consommées (en g ou ml) pour les 110 aliments, 12 boissons non alcoolisées et 4 boissons alcoolisées (qtte1j → qtte136j)
 - Apport en nutriment, alcool compris, pour les 43 nutriments considérés (totv1 → totv43)
 - Apport en nutriment, hors alcool, pour les 43 nutriments considérés (tot_sfalc_v1 → tot_sfalc_v43)

Pour sélectionner les sujets avec un questionnaire alimentaire valide, prendre uniquement les sujets avec selectionFFQ=1.

Les statistiques de base (N, moyenne, écart-type, min, max), pour les sujets ayant un questionnaire alimentaire valide, sont présentées dans le fichier Excel Descriptif_AlimMere_Jan2015.xlsx.

PONDERATION SPECIFIQUE

L'auto-questionnaire en maternité fait l'objet d'une pondération (pondaAQM) définie sur les questionnaires exploitables du point de vue de l'ensemble des thèmes abordés. La documentation de cette pondération présente en détail les caractéristiques de cet échantillon de répondants. Les analyses précédentes, explicitées ci-dessus, permettent une autre définition de l'échantillon exploitable (selectionFFQ), cette fois du seul point de vue du thème de l'alimentation et spécifiquement adaptée aux analyses prévues. Une demande a été faite à l'équipe Elfe pour qu'une pondération (pondalimselectFFQ) soit réalisée sur la base de cet échantillon.

Cette pondération permet l'inférence à la population cible Elfe. Pour rappel, la population d'inférence ELFE est celle des nourrissons vivants, nés en 2011 en maternité métropolitaine, issus d'un accouchement au plus gémellaire, hors grands prématurés, ayant une mère majeure, en mesure de donner un consentement éclairé notamment dans l'une des langues proposées (français, anglais, arabe ou turc) et dont les parents ne résidaient pas temporairement en métropole.

La pondération proposée pondalimselectFFQ repose sur la pondération maternité M00E_PONDVALC2 multipliée par un coefficient corrigeant la phase de non réponse. L'unité statistique est le nourrisson. Les caractéristiques de la pondération maternité, ainsi que la justification des choix techniques de repondération peuvent être retrouvées dans les documentations des pondérations *maternité* et *deux mois*.

L'échantillon sélectionné contient 14 545 enfants, l'échantillon en maternité contient 18 329 - 123 = 18 206 enfants (06/2015). 123 enfants n'ont pas de pondération maternité soit parce

qu'ils sont en dehors des 25 jours d'enquêtes (71 d'entre eux) soit parce qu'ils ont demandé la suppression de leur données. Ainsi, l'échantillon final des répondants sélectionnés disposant d'une pondération est en réalité de 14 485.

Le coefficient de redressement de la non réponses est calculé par la méthode des groupes de réponses homogènes (**GRH**). Le modèle statistique du mécanisme de non réponse contient les facteurs suivant : *Région d'habitation, âge de la mère, type d'alimentation du nourrisson (allaitement, lait premier âge, mixte) , situation professionnelle de la mère au moment de la grossesse, csp de la mère, situation de couple (en couple ou pas), la mère a suivi des séances de préparation à la naissance, la mère est partie en vacances pendant la grossesse, statut immigré du ménage, niveau d'étude de la mère, IMC de la mère.*

A partir des scores triés résultants de cette régression, 40 groupes de réponses homogènes ont été créés et pondérés par la pondération en maternité. Les répondants représentent les non répondants au sein de chaque groupes ce qui détermine un coefficient de **repondération** pour chaque répondant.

Un **calage** est effectué sur les marges de l'ENP 2010 et de l'État Civil 2011 à partir des variables suivantes : *Age, Région, Statut immigré, État matrimonial, Primiparité, Niveau d'étude.*

Les données manquantes (à très petit taux) pour ces variables ont été imputées par hotdeck aléatoire (avec tirage proportionnel au poids en maternité) afin que ces individus ne soient pas exclus du calage. Le calage est réalisé par la méthode du raking ratio implémentée dans CALMAR (INSEE).

Dans la recherche d'un équilibre entre le biais et la variance, les poids sont tronqués à 250. Enfin la somme des poids est ajustée pour extrapoler à la population cible estimée à 764 000 enfants (nés en 2011, en France métropolitaine, d'un accouchement au plus gémellaire, hors grand prématuré, sous critère d'éligibilité). En moyenne le poids d'un enfant de l'échantillon est de 52. Le détail des variables utilisées peut être retrouvé dans la documentation de la pondération PondaQM sur l'ensemble de l'auto-questionnaire en maternité.

RÉFÉRENCES

Black A.E. (2000) Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. *International Journal of Obesity and Related Metabolic Disorders* **24**, 1119-1130.

Goldberg G.R., Black A.E., Jebb S.A., Cole T.J., Murgatroyd P.R., Coward W.A., et al. (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *European Journal of Clinical Nutrition* **45**, 569-581.

Herberg S. (2006) *Table de composition des aliments SU.VI.MAX*. Economica.

Herberg S., Deheeger M. & Preziosi P. (2002) *SU.VI.MAX. Portions alimentaires : Manuel photos pour l'estimation des quantités*. Editions Polytechnica, Paris.