Eur J Nutr 42 : 67 – 72 (2003)
DOI 10.1007/s00394-003-0412-8

ORIGINAL CONTRIBUTION

Matthijs Dekker
Ruud Verkerk

Dealing with variability in food production chains: a tool to enhance the sensitivity of epidemiological studies on phytochemicals

Summary *Background* Many epidemiological studies have tried to associate the intake of certain food products with a reduced risk for certain diseases. Results of these studies are often ambiguous, conflicting, or show very large deviations of trends. Nevertheless, a clear and often reproduced inverse association is observed between total vegetable and fruit consumption and cancer risk. Examples of components that have been indicated to have a potential protective effect in food and vegetables include antioxidants, allium compounds and glucosinolates. *Aim* The food production chain can give a considerable variation in the level of bioactive components in the products that are consumed. In this paper the effects of this variability in levels of phytochemicals in food products on the sensitivity of epidemiological studies are assessed. *Methods* Information on the effect of variation in different steps of the food production chain of *Brassica* vegetables on their glucosinolate content is used to estimate the distributions in the levels in the final product that is consumed. Monte Carlo simulations of an epidemiological cohort study with 30,000 people have been used to assess the likelihood of finding significant associations between food product intake and reduced cancer risk. *Results* By using the Monte Carlo simulation approach, it was shown that if information on the way of preparation of the products by the consumer was quantified, the statistical power of the study could at least be doubled. The statistical power could be increased by at least a factor of five if all variation of the food production chain could be accounted for. *Conclusions* Variability in the level of protective components arising from the complete food production chain can be a major disturbing factor in the identification of associations between food intake and reduced risk for cancer. Monte Carlo simulation of the effect of the food production chain on epidemiological cohort studies has identified possible improvements in the set up of such studies. The actual effectiveness of food compounds already identified as cancer-protective by current imprecise methods is likely to be much greater than estimated at present.

Key words Monte Carlo simulations – epidemiology – food production chain – processing – sensitivity – glucosinolates

Received: 15 July 2002
Accepted: 26 January 2003

Matthijs Dekker (✉) · R. Verkerk
Wageningen University
Product Design and Quality Management Group
Dept. of Agrotechnology and Food Sciences
P. O. Box 8129
6700 EV Wageningen, The Netherlands
E-Mail: matthijs.dekker@wur.nl

Introduction

Many epidemiological studies have indicated the protective role of vegetable and fruit consumption for various diseases. Epidemiological data considering the effect of vegetable and fruit intake on cancer risk have been reviewed by several investigators [1–3].

Although there is clearly a positive trend in the protective role of vegetables and fruit for various cancers, the outcomes of individual epidemiological studies sometimes conflict. This is clearly illustrated in a review by Steinmetz and Potter [2], showing statistically significant inverse associations for 15 of the 21 examined case-control studies between one or more vegetable/fruit categories and colon cancer. Four of the studies did not show significant inverse associations for any vegetable/fruit category and in two studies statistical sig-

EJN 412

nificance was not reported. Many sources of uncertainty and variation can be responsible for these discrepancies.

Accurate assessment of dietary intake is of importance to investigate associations between vegetable/fruit consumption and cancer incidence. Several methods are available to assess intake, but they all have their specific measurement errors. Dietary assessment methods can be roughly divided into methods assessing current diet, consisting of records and 24h recalls and methods assessing habitual diet, consisting of dietary history and food frequency questionnaires. The intake data derived by these methods combined with analytic data from food composition tables or from chemical analysis, result in individual nutrient (and non-nutrient) intakes. In addition to these methods, some biological indicators (biomarkers) of dietary exposure have been developed.

Food composition tables are important tools for epidemiological studies on nutrients and non-nutrient phytochemicals in relation with diseases. However, accurate knowledge of the nutrient and non-nutrient intake of individuals and groups of people requires information on the contents of prepared foods, in other words prior to consumption. Unfortunately, dietary calculations are frequently made on the basis of foods as brought into the kitchen. Also, for many components of potential interest for health protection (like flavonoids and other polyphenols, lignans, carotenoids, glucosinolates, peptides, lactic acid bacteria) no or only limited food composition data are available [4–6].

It is demonstrated that many steps in the food production chain, like cultivation, storage, processing and preparation of the vegetables (Figs. 1 and 2), can have a large impact on levels and thus intake of phytochemicals [7]. This variation has been investigated for the group of glucosinolates from Brassica vegetables but is likely to occur for most classes of health-promoting phytochemicals. The potential health benefits of glucosinolates have been reviewed extensively in [8]. Epidemiological studies on Brassica vegetables are given in [3, 9, 10].

Epidemiology often does not take into account the effects of food processing and preparation on glucosinolate content, and thus of the consequences for potential health-protection. It can be hypothesized that the high variable levels of glucosinolate intake explain the inconsistent epidemiological findings. It is impossible to systematically gather information on all possible variations in all steps of the production chain. Therefore, a predictive modeling approach was developed to estimate the effects of variation in conditions and processes on the level of glucosinolates and breakdown products [7, 11, 12].

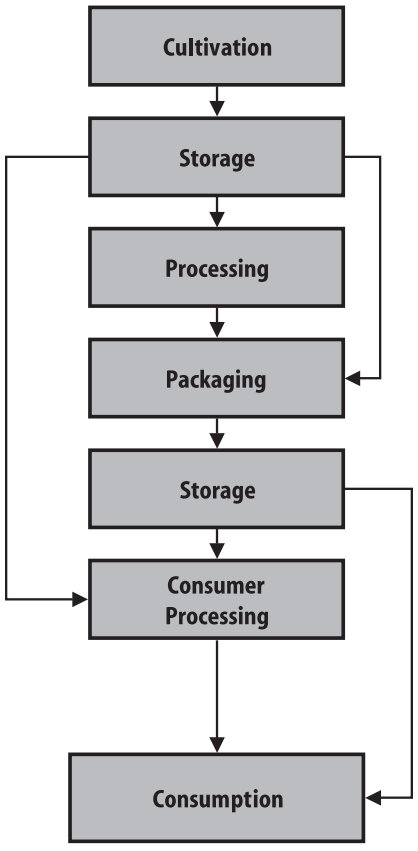

Fig. 1 Schematic representation of the food production chain of *Brassica* vegetables

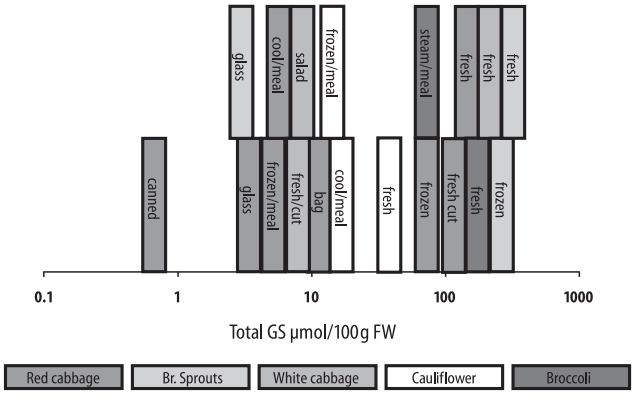

Fig. 2 Distribution in the content of total glucosinolates of different *Brassica* products (data from [11])

Simulation strategy

The steps in the simulation experiments of the epidemiological cohort studies are schematically represented in Fig. 3. The hypothesis of the study is that whether an individual will develop cancer depends on his/her absolute risk of cancer which depends in part on the intake

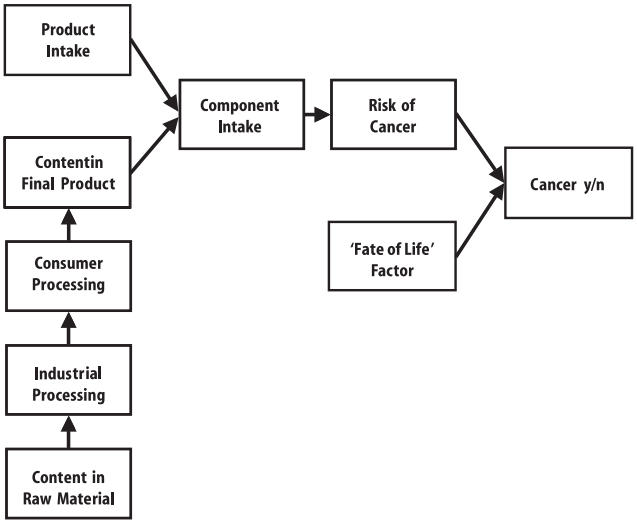

Fig. 3 Schematic representation of the applied strategy for the simulation of epidemiological cohort studies on the protective effect of foods on cancer incidence

of protective compounds from foods. The simulation strategy as depicted in Fig. 3 will give information as to whether or not the cohort study will give a significant association between product intake and the risk on developing cancer.

Because of the availability of information of the distribution of the levels of glucosinolates and of some important effects of the food production chain, these compounds have been taken as the protective compounds.

Three consumer groups were simulated with low (0–200), medium (200–400) and high (400–1000 g/week) *Brassica* vegetable intakes. In total 30,000 consumers were simulated with a distribution in intakes as shown in Fig. 4.

Distributions of the glucosinolate content in raw materials and of the effect of industrial processing, storage and consumer processing were estimated from published experimental data for *Brassica* vegetables [7, 11, 12]. The distributions were described by log-normal distributions:

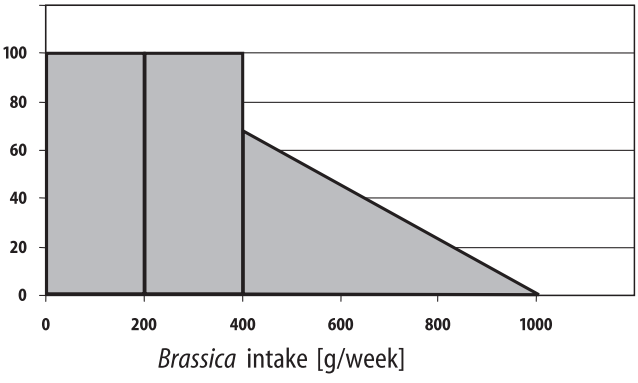

Fig. 4 Distribution of the 30,000 individuals in the three *Brassica* intake groups

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma'}} e^{-\frac{1}{2}\left[\frac{\ln(x)-\mu'}{\sigma'}\right]^2} \qquad (1)$$

with

$$\mu' = \ln\left[\frac{\mu^2}{\sqrt{\sigma^2+\mu^2}}\right] \quad \text{and} \quad \sigma' = \frac{1}{2}\ln\left[1+\left(\frac{\mu}{\sigma}\right)^2\right]$$

where x, μ and σ are continuous parameters. The distributions were truncated ($0 < x \leqq \max$), the parameters used for the variation at the different steps of the production chain are given in Table 1. The values for cultivation give a distribution in level in the raw material (in μmole/100 g fresh weight).

The resulting variation in the glucosinolate content in the products that are consumed were calculated by multiplying each randomly picked level of raw material with the randomly picked industrial and consumer processing effect. The component intake was calculated by multiplying the sampled product intake with the calculated (as described above) glucosinolate content in the product.

To calculate the relative risk of each consumer, a relation was assumed between the intake of a health-promoting component in the diet (glucosinolates in this study) and the reduced risk for the development of cancer. This dose-response relation was described by the equation:

$$RR = (1 - MR)e^{\frac{-EF \cdot I}{100}} + MR \qquad (2)$$

in which *RR* is the relative risk for cancer development, *RM* is the minimum risk, *EF* is the effectiveness of the component and *I* is the intake of the component. To assess the statistical power of epidemiological studies, different effectiveness factors (EF) for protecting against cancer ($EF = 1$–20) and a fixed minimum risk ($MR = 0.2$) were taken (Fig. 5).

For each simulated consumer the relative risk was transformed in an absolute risk value by multiplying the RR by 0.01 (this value is estimated from the reported

Table 1 Distribution curves and parameters used for the three steps in the production chain taken into account in the Monte Carlo simulations (μ mean; σ standard deviation)

Step in chain	Type of distribution	μ	σ	max.
Cultivation	Truncated log normal	100	100	1000
Industrial processing	Truncated log normal	0.7	0.6	3
Domestic cooking	Truncated log normal	0.5	0.4	1

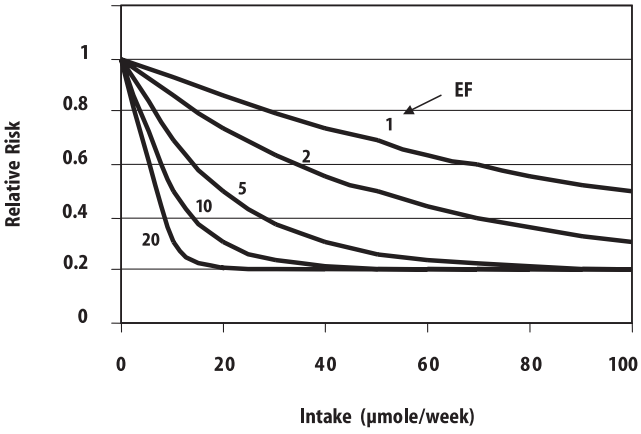

Fig. 5 Relative Risk (RR) for developing cancer as a function of the glucosinolate intake, as calculated with Eq. (2) for five different values of the Effectiveness Factor (EF) as shown within the graph

cases of colon/rectal cancer in a recent cohort study [10]). In order to simulate whether an individual will develop cancer this calculated absolute risk was compared with a 'fate of life' factor that was randomly picked from a uniform distribution between 0 and 1. If this 'fate of life' factor was lower than the absolute risk value the consumer was marked as a cancer patient in the study.

All Monte Carlo simulations were performed using Pallisade @RISK software as an add-in for Microsoft Excel.

Simulation results and discussion

From the three product intake groups (0–200, 200–400, 400–1000 gram *Brassica* vegetables per week), the intake of glucosinolates was calculated by taking into account all the variation that could be due to three steps in the production chain (cultivar, industrial processing and domestic cooking). The resulting compound intake distributions are shown in Fig. 6. A considerable overlap in compound intake levels can be seen between the groups, owing to the large possible variation caused by the production chain. In reality some of the variation caused by

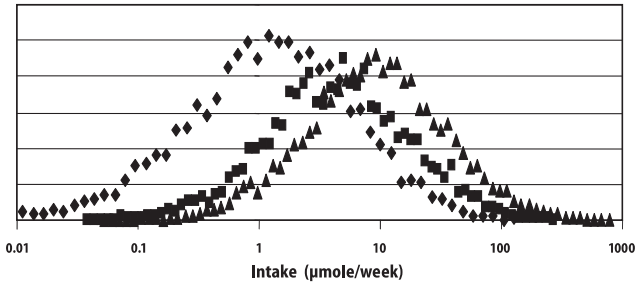

Fig. 6 Calculated glucosinolate intake of the three *Brassica* intake groups (◆: 0–200, ■: 200–400, ▲: 400–1000 g *Brassica*/week)

cultivar differences will level out because a range of different cultivars will be consumed over the years that these studies take. The type of products and the way of domestic cooking, however, can be expected to be not so variable for each individual and therefore the variation caused by these steps will not level out. Because of the fact that only three sources for variation resulting from the food production chain were taken into account, while the complete chain will consist of at least 6 steps (Fig. 1) with each multiple sources of variation [7], the real variation might in fact be underestimated by the approach taken here.

By using the compound distributions as shown in Fig. 6 the risk of cancer of each individual was calculated. By the strategy described above this absolute risk could be converted to the number of cancer patients in the three product intake groups in the cohort.

This assessment was done for three scenarios:

A. The ideal situation, in which the production chain causes no variation (median values of glucosinolate content and processing effects were used, as given in Table 1).
B. The 'real' situation, in which the three steps that are considered in the production chain give the variation as can be expected from experimental observations (the whole distributions as given in Table 1 were taken into account).
C. An 'intermediate' situation, in which the first two steps of the production chain give the expected variation, but the domestic cooking effect is assumed to be known for each individual, and therefore no additional variation is introduced by this step.

The resulting relative risks for the three product intake groups is given in Fig. 7 for a value of the effectiveness factor (EF) of five. As can be seen in Fig. 7A, the health protecting effect of *Brassica* vegetables could be very significantly assessed if it were the case that the food production chain gave products with a constant level of glucosinolates. For the 'real' situation, however, in which the production chain gives an enormous variation in the levels of glucosinolates in products, no health protective effect could be seen from the cohort study (Fig. 7B). So even though the glucosinolates present in *Brassica* vegetables have quite a potent health protective effect (see Fig. 5 for EF = 5), this protective effect cannot be seen in a large cohort study where only the product intake is assessed as an input parameter. If the study design were to include information on the cooking behavior of the individuals the unknown variation in the levels of glucosinolates in the consumed products could be reduced. In this case the cohort study would conclude a significant health protective effect ($p < 0.05$) of the consumption of *Brassica* vegetables on the development of cancer (Fig. 7C).

To establish the improvement in the statistical power of cohort studies, the simulations were performed with

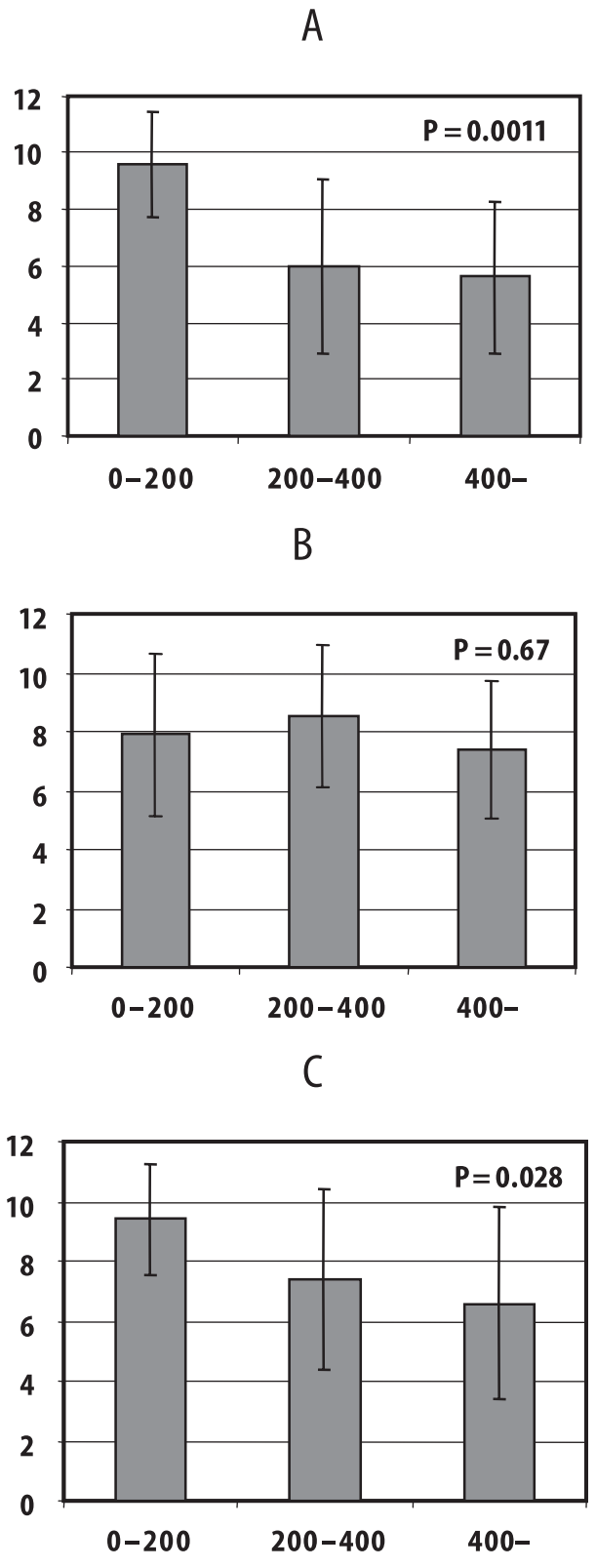

Fig. 7 The relative risk for cancer for the three *Brassica* intake groups as observed in the simulated cohort studies with three different scenarios: **A**, **B** and **C**. See text for explanation of the scenarios

values of EF between 1 and 20. The significance of the resulting health protecting effect is shown in Fig. 8, in which the p-value between the relative risks of the high and low consumption groups is plotted versus the EF value. From this figure it can be concluded that the statistical power can be at least doubled by incorporating information of cooking habits in the design of the study. This can be seen in Fig. 8 by looking at the EF value for which the study design gives a statistical significant difference in the relative risks. If the design were to allow for assessment of all the possible variation of the food production chain, the statistical power could be increased by at least a factor of five.

In the design of epidemiological studies, the information that can be gathered from the individual on their cooking habits will be limited by practical constraints like the number of questions in a questionnaire or diary and the accuracy of individuals in describing their habits. However, in combination with a predictive modeling approach as described by Dekker et al. [7] and Verkerk et al. [11], the quantification of cooking effects on health parameters can be made in a computerized way, using cooking time and the ratio of water to vegetable as inputs.

This simulation study clearly shows that, by improving the design of epidemiological studies, health protective compounds in foods can be identified with more statistical power. In addition, a simulation approach as presented here can also be used to establish the minimum size of the cohort to be used in an epidemiological study. Another important conclusion is that the underlying protective effect of certain compounds in foods will be much larger than is shown in the present epidemiological studies that identified a significant protection by the intake of certain food product categories.

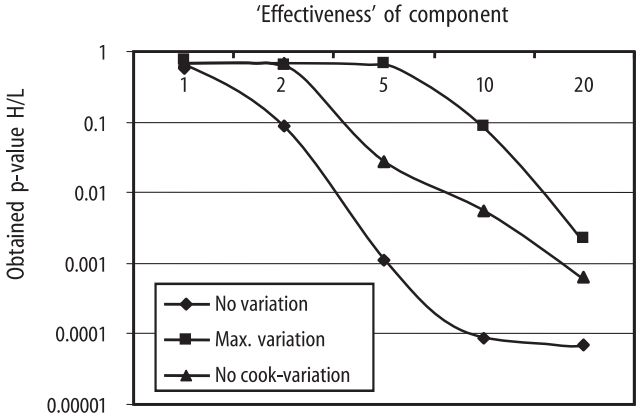

Fig. 8 Significance of the observed protective effect of *Brassica* vegetables on cancer incidence as a function of the Effectiveness Factor of glucosinolates for the three study scenarios as explained in the text

References

1. Steinmetz KA, Potter JD (1991) Vegetables, fruit and cancer. I. Epidemiology. Cancer Causes and Control 2:325–357
2. Steinmetz KA, Potter JD (1996) Vegetables, fruit and cancer prevention: a review. J Am Diet Assoc 96:1027–1039
3. Verhoeven DTH, Goldbohm RA, van Poppel G, Verhagen H, van den Brandt PA (1996) Epidemiological studies on Brassica vegetables and cancer risk. Cancer Epidemiology, Biomarkers & Prevention 5:733–748
4. Hertog MGL, Hollman PCH, Katan MB (1992) Content of potentially anticarcinogenic flavonoids of 28 vegetables and 9 fruits commonly consumed in the Netherlands. J Agric Food Chem 40: 2379–2383
5. Hertog MGL, Hollman PCH, van de Putte B (1993) Content of potentially anticarcinogenic flavonoids of tea infusions, wines and fruit juices. J Agric Food Chem 41:1242–1246
6. Hollman PCH, Arts ICW (2000) Flavonols, flavones and flavanols-nature, occurrence and dietary burden. J Sci Food Agric 80:1081–1093
7. Dekker M, Verkerk R, Jongen WMF (2000) Predictive modelling of health aspects in the food production chain: a case study on glucosinolates in cabbage. Trends in Food Science and Technology 11:174–181
8. Mithen, RF, Dekker, M, Verkerk, R, Rabot S, Johnson, IT (2000) The nutritional significance, biosynthesis and bioavailability of glucosinolates in human foods. J Sci Fd Agr 80:967–984
9. Michaud DS, Spiegelman D, Clinton SK, Rimm EB, Willett WC, Giovannucci EL (1999) Fruit and vegetable intake and incidence of bladder cancer in a male prospective cohort. J Natl Cancer Inst (Bethesda) 91:605–613
10. Voorrips LE, Goldbohm RA, van Poppel G, Sturmans F, Hermus RJJ, van den Brandt PA (2000) Vegetable and fruit consumption and risks of colon an rectal cancer in a prospective cohort study. The Netherlands cohort study on diet and cancer. American Journal of Epidemiology 152 (11):1081–1092
11. Verkerk R (2002) Evaluation of glucosinolate levels throughout the production chain of Brassica vegetables; towards a novel predictive modelling approach. PhD Thesis, Wageningen University, The Netherlands
12. Verkerk R, Dekker M, Jongen WMF (2001), Post-harvest increase of indolyl glucosinolates in response to chopping and storage of Brassica vegetables. J Sci Fd Agric 81:953–958