Chapter 11
Agriculture, Food and Nutrition Security: Concept, Datasets and Opportunities for Computational Social Science Applications

T. S. Amjath-Babu, Santiago Lopez Riadura, and Timothy J. Krupnik

Abstract Ensuring food and nutritional security requires effective policy actions that consider the multitude of direct and indirect drivers. The limitations of data and tools to unravel complex impact pathways to nutritional outcomes have constrained efficient policy actions in both developed and developing countries. Novel digital data sources and innovations in computational social science have resulted in new opportunities for understanding complex challenges and deriving policy outcomes. The current chapter discusses the major issues in the agriculture and nutrition data interface and provides a conceptual overview of analytical possibilities for deriving policy insights. The chapter also discusses emerging digital data sources, modelling approaches, machine learning and deep learning techniques that can potentially revolutionize the analysis and interpretation of nutritional outcomes in relation to food production, supply chains, food environment, individual behaviour and external drivers. An integrated data platform for digital diet data and nutritional information is required for realizing the presented possibilities.

11.1 Introduction

The global goal of ending hunger and malnutrition (Sustainable Development Goal-2) by 2030 is off track as the numbers of food insecure and malnourished people are increasing (Fanzo et al., 2020). The number of undernourished people climbed to 768 million in 2020 from 650 million in 2019 (FAO, 2021), belonging mainly to the Asian (>50%) and African continents (25%). This might be further increased in...
the context of the economic disruption caused by COVID-19 pandemic and global price hikes due to recent Russia-Ukraine conflict.

Some may consider it ironic that a large proportion of the undernourished people, who cannot afford healthy diets, are those involved in the food production, including subsistence farmers and farm labourers (Fanzo et al., 2022). In addition, low- and middle-income (LMIC) as well as wealthy countries are burdened with overweight (BMI > 25), obesity (BMI > 30) and diet-related non-communicable diseases (Ferretti & Mariani, 2017; Global Panel on Agriculture and Food Systems for Nutrition, 2016). As such, there are increasing calls for agricultural and food system innovations and policies that can enhance diets and improve availability of quality foods for better nutrition and health outcomes (Fanzo et al., 2022; Global Panel on Agriculture and Food Systems for Nutrition, 2016).

Nevertheless, the linkages of global and national food systems to nutritional outcomes are complex and are influenced by diverse macro-level (trade, market access, climate change, technology, conflicts, wealth distribution, agricultural policies, etc.) and micro- and meso-level factors (farm types, income, gender considerations, diet preferences, attitude and beliefs, inter- and intra-household dynamics and power, cooking methods, sanitation, among others). A deeper understanding of these multi-scale (micro-, meso- and macro-level) drivers of nutritional outcomes is vital in devising agricultural policies and programmes and hence transforming the agri-food sector to meet the goal of ending hunger and malnutrition (Global Panel on Agriculture and Food Systems for Nutrition, 2016). There is a wide recognition of inadequate methods, data and metrics for understanding agri-food systems relationships to nutritional outcomes and dynamics (Marshall et al., 2021; Micha et al., 2018; Sparling et al., 2021). Towards this end, new sources of data and emerging computational social science methods may offer possibilities to test novel conceptual frameworks as well as empirical and experimental examination of the complex relationships and pathways. The current chapter focuses on how the availability of digital data and computational social science methods can support modelling and the analytics of a complex portfolio of factors (and their interactions) influencing food and nutritional outcomes. It also highlights the need of data-sharing protocols and platforms for fully utilizing the potential of emerging data and analytical tools for generating meaningful policy insights (Müller et al., 2020; Takeshima et al., 2020).

11.2 The Complex Pathways to Nutritional Outcomes: A Conceptual Note

Agricultural production and consequently nutrient availability and consumption are interrelated through complex pathways span across spatial and time scales. Nutritional outcomes (Sparling et al., 2021) are driven by a range of factors including food production, consumer purchasing power, trade and market systems as
well as food transformation and consumer behaviour (Global Panel on Agriculture and Food Systems for Nutrition, 2016). The downturns in economies, climate stress and conflicts can contribute to changes in consumption practices that lead to malnutrition, while trade policies and supply chain infrastructure can impact food prices that influence the costs of food necessary for healthy diets that in part drive nutritional outcomes (FAO, 2021). Climate variability and extreme events can lead to losses in agricultural production and increased import demand from affected countries, leading to food price volatility (Willenbockel, 2012; Chatzopoulos et al., 2020). Recession or reduction in economic activity at country level can also lead to unemployment and reduction in wages and income, which may force households to shift to energy dense and cheaper foods (including ‘junk food’) instead of purchasing and consuming nutritious foods (Dave & Kelly, 2012). Income and social inequality amplify the impact of climate stresses or economic downturns in terms of access to nutritious diets (FAO, 2021). Lower productivities and low efficiency of supply chains can also lead to higher prices for diverse food groups needed for healthy diets. Conflicts or health crises like the COVID-19 pandemic disrupt the movement of goods, increase in prices of healthy foods and decrease in their availability (Amjath-Babu et al., 2020). Conflicts can also reduce access to capital, energy, labour or land and hence impact food production (FAO, 2021).

In the case of farm households, the raw nutrient availability for a household is determined by own production used for self-consumption, purchased food from market using the farm household income and food received through informal exchanges and social safety nets. Farm household income is determined by the yield (sold in market) of various farm enterprises (cereal crops, vegetables, cash crops, livestock, aquaculture, etc.), their farmgate price levels and the cost (inputs including land, labour, machinery, fertilizers, pest control, etc.) of production in addition to any available rental income, off-farm income and remittances. Farm production and income is further conditioned by environmental stresses and the state of natural resources (e.g. soil, water), agricultural policies and market infrastructure. Apart from these drivers, the technology available at farm level influences yield performance and post-harvest losses that impact food availability and access (Müller et al., 2020). The direct (self-consumption) and indirect (as a source of income for market purchase of food consumed) role of farm production in food nutrient availability depends on the strength and quality of market linkages (Bellon et al., 2020; Sibhatu et al., 2015).

The net energy and macro- and micronutrient availability to men, women and children are further conditioned by diet preferences, cooking methods, gender norms, nutritional knowledge, attitudes and beliefs (Monterrosa et al., 2020). A deficit in net availability and effective consumption of nutrients compared to requirement of individual members can potentially lead to malnutrition that can manifests in stunting, wasting of children, nutrition-deficit disorders as well as nutrition-related non-communicable diseases (N-NCD) in children and adults. Stunting, wasting and N-NCD are conditioned not only by nutritional deficiencies but also by nutrient utilization (determining bioavailability of nutrients through metabolic pathways) capacity of human bodies (Millward, 2017) and sanitary,
hygiene and water quality conditions. Conversely overconsumption of high-calorie, low-nutrient foods can lead to overweight and obesity (Astrup & Bügel, 2019). Women’s empowerment in terms of time, income and asset control can also positively influence the nutritional and health-related outcomes (Herforth & Ballard, 2016). The agriculture-nutrition pathway map by Kadiyala et al. (2014) includes health-care expenditure, health status as well as women’s employment as additional determinants for nutritional outcomes. Figure 11.1 provides a comprehensive overview of the complex path of nutritional outcomes.

In the case of affordability of diets, nutrient adequate (e.g. the advised ‘EAT-Lancet diet’) diets is not affordable for 1.5 billion poor people globally (Hirvonen et al., 2020). Even in European Union, around 10% of population in 16 countries faces financial issues in affording healthy diets (Penne & Goedemé, 2021). Nutrient-rich food items are often costly to grow, store and transport compared to starchy food. Oil and sugar tend also to have a longer shelf life and is easier to transport (Fanzo et al., 2022). This calls for also further understanding on ways to make the nutritious food more affordable as high prices of nutritious food or its volatility can negatively impact consumption among the poor. Conversely, lower prices of sugar and sugar-rich food prices are related to higher prevalence of overweight and obesity (Headey & Alderman, 2019). These point to the importance of a better understanding of the macro-economic policies on nutritional outcomes as well as disconnects between agriculture-nutrition pathways.

The discussion so far highlights the requirement of deeper understanding the complex pathways linking nutritional security, health outcomes and public policies, especially for the most vulnerable groups (children and women). Below, we discuss existing modelling-based approaches as well as the role of emerging digital data and computational methods in opening new frontiers in quantifying the ex ante impacts of regional or national food and nutrition policies by unravelling the complex interactions of the macro-meso-micro-factors.

### 11.3 Current Ex Ante Analytical Models for Nutritional Policy Insights

In case of existing ex ante assessments (nutritional outcomes of agricultural policy), three studies are discussed here for documenting the current state of the art of methods employed. Lopez-Ridaura et al. (2018) took a nutrient-balancing approach where self-consumed farm products and the net annual farm income derived from all farm enterprises were converted to energy equivalents and compared with annual food energy requirement of households. Although the model focused on calories, the simplified relation allowed simulations of yield changes due to new technologies and their impact on potential household-level food availability ratios. The study provides a framework that could also be extended to macro and micronutrient availability and consumption (Bizimana & Richardson, 2019).
Fig. 11.1 Macro-level drivers and household-level nutrition outcome pathways of farm households [expanded framework based on an initial framework presented in Lopez-Ridaura et al. (2018)]
FARMSIM model is able to simulate the impact of net cash income from all farm enterprises on consumption of nutrients such as protein, calories, fat, calcium, iron and vitamin A. To represent the complex interrelations of macro- and micro-factors affecting food production and nutritional outcomes, current modelling approaches addressing nutritional security questions often use local-level proxies of macro-determinants (e.g. yield functions) or through representations of key variables (e.g. food prices or farm sizes) (Müller et al., 2020). In the case of the FARMSIM model, market prices are simulated (using Monte Carlo approaches) using probability distributions obtained from historical data or expert opinion, while yield distributions are generated by crop yields generated by the APEX (Agricultural Policy/Environmental eXtender) simulation model using historical weather data and plant growth parameters. These are matched to different technologies considered in simulation. Consideration of stochastic prices and yields allows modelling risk behaviour using stochastic efficiency with respect to a function (SERF).

FSSIM-Dev (European Commission. Joint Research Centre, 2020) is a farm-level model based on positive mathematical programming (PMP) which does not consider risk as the model is deterministic. FSSIM-Dev considers the non-separability of production and consumption decisions of farming households: It maximizes the utility from both the production and the consumption of food, and the decision to rely on home production or to go the market is governed by transaction costs. The model considers annual income beyond farm income by including subsidies, pensions, off farm income, remittances and other transfers as exogenous variables. Farm income is linked to the consumption by a linear expenditure function of uncompressible consumption below which consumption may not fall and supernumerary consumption, which is modelled as a fixed proportion (marginal budget share) of net income. FSSIM-Dev model is capable to generate food and nutrition security indicators as carbohydrates, proteins and lipids from the simulated food consumption. The simulation of micronutrients is not yet attempted by the existing model. Figure 11.1 shows an extended conceptual modelling frame that can offer wider insights to the questions related to nutritional impacts of agricultural policies, food environment, sanitation, etc.

The quoted policy simulation studies analysed policy impact on availability of macronutrients such as carbohydrates or proteins and had limited capacity in dealing with micronutrients. In addition, modelling efforts currently have limited ability to consider the access (income levels, impact of social safety nets, informal exchanges of food) and stability (seasonality and occasional shocks), gender roles (intra-household food allocation, women empowerment) and utilization (bioavailability) dimensions of the nutrient security question. Net nutritional availability is also affected by cooking methods, knowledge, attitude and beliefs that are not always integrated in modelling exercises to reduce complexity. In case of availability of nutrients for a given individual, distribution of food within households adds another layer of complexity. Despite the fact that ensuring adequate nutrition at an individual’s level is at the heart of nutritional challenges, policy insights at this level are generally lacking. New sources of digital data and computational methods are expected to address the stated challenges.
11.4 The Data Scarcity for Nutritional Modelling and Analytics

The scarcity, within countries and among countries for harmonized data on food consumption and nutrition, is a major challenge for initiatives aimed at addressing global nutritional challenges. Currently, major data sources used for analytics are the specialized household surveys [demographic health surveys (DHS), multi-indicator cluster surveys (MICS), dietary intake surveys, consumption expenditure surveys, Living Standard Measurement Studies (LSMS), Food Security Monitoring System (FSMS) etc.]. These tend to include both economic-related variables and detailed diet data (Buckland et al., 2020). There is a need for efforts to make the datasets (cleaned data on food consumption and their nutritional equivalents) open through platforms with defined standards and efficient infrastructure to share data (de Beer, 2016). To make the open data sharing a reality, technological, legal and ethical challenges need to be considered. Communities like smallholder farmers hold data that, when subject to analytics, can be used to improve their well-being. But there is an absence of platforms and mechanisms that enable rapid and regular data acquisition and sharing (de Beer, 2016). Traka et al. (2020) suggested making food and nutrition data FAIR (findable, accessible, interoperable and reusable).

An integrated data perspective of agriculture-food, nutrition and health is required for meaningful interventions ensuring sustainable production, shift in diets and reduction in non-communicable nutrition-related diseases (Traka et al., 2020). Even if data are available, they are often not available at sub-national levels, or may not be disaggregated across demographic groups, or may be out of date, in addition to a range of additional data quality challenges. Key complementary information required in preparing such databases is country-specific food composition tables (FCTs) that can be used to convert food products to their nutrient value, although often these tables are not comprehensive or may not even be available. As such, here is a need of coordinated effort to make sure that comprehensive FCTs are available (Ene-Obong et al., 2019).

The distribution of food within households adds another layer of difficulty, since information regarding intra-household distribution is lacking in many surveys. The disaggregation of household consumption data using adult male equivalent (AME) weights is a relevant disaggregation method (Coates et al., 2017) and is based on household members’ relative caloric requirement. This deterministic method can be improved by adding error observed in dietary energy expenditure prediction models. Côté et al. (2022) compared traditional regression models against machine learning models in predicting individual vegetable and fruit consumption and did not observe a major improvement. Nevertheless, the scope of using machine learning or other innovative methods in disaggregation of consumption data is currently underexplored. Lager sets of food consumption data, disaggregated among household members, are required for validating and fine-tuning the methods.
11.5 Novel Digital Food and Nutrition Data for Computational Analytics

Global Individual Food consumption data Tool (GIFT-FAO/WHO) initiative aims to make harmonized data freely accessible online through an interactive web platform. This tool is based on FoodEx2, which is a food classification and description protocol (food items are coded with distinct hierarchy) developed by the European Food Safety Authority (EFSA) to standardize the 24 h recall data (Leclercq et al., 2019). There are also increasing attempts to collect nutritional data using telephonic surveys. Lamanna et al. (2019) reports that efforts to collect nutrition data from rural women in Kenya through telephone surveys result in a 0–25% increase in nutritional scores [minimum dietary diversity for women (MDD-W) and minimum acceptable diet for Infants and young children (MAD) estimates compared to face-to-face interviews]. This points to the potential of using digital tools and methods for data collection.

Nutrition apps for collecting diet information are increasingly available for diet recording and monitoring (Campbell & Porter, 2015). Hundreds of nutrition-related mobile apps are available, but their utility in tracking healthy food consumption and nutrition data generation is still limited. Fallaize et al. (2019) compared popular nutrition-related apps (Samsung Health, MyFitnessPal, FatSecret, Noom Coach and Lose It!) that assess macronutrients and micronutrients as well as energy from consumed food, against a reference method. They showed that apps are in general capable to assess macronutrient availability, while micronutrients estimates were inconsistent. Another similar study on nutritional apps (FatSecret, YAZIO, Fitatu, MyFitnessPal and Dine4Fit) showed inconsistent results on macronutrients and energy intake estimates (Bzikowska-Jura et al., 2021).

The efforts to use digital diet information collected by these or similar apps are so far very limited (Martinon et al., 2022), although they offer large potential. Smartphone image-based automatic food recognition and dietary assessment tools are currently emerging. These tools are attempting to identify, classify and estimate volume of food intake and nutrient content estimation. Machine learning and deep learning approaches are also now being used for classifying food items in a meal, which depends on generic and comprehensive food image datasets for training data. Deep learning approaches including convolutional neural network have been suggested as being more effective than machine learning algorithms such as support vector machines (SVM) and K-Nearest Neighbor (KNN) for this purpose due to their higher classification efficiency (Ciocca et al., 2020). Nevertheless, the quantification of the mass of food by visual assessment of volume and density is much more challenging. The estimation of calories and nutrients can be error prone due to poor classification and mass estimations. Further research and development of fully automated nutritional content detection applications using smartphones may transform it to a game changer for nutritional data (Subhi et al., 2019). An assessment of the meal snapp app by Keeney et al. (2016) showed that calorific values generated by such apps are comparable to standard application (Nutritionist
Pro™ used by dietitians). Nevertheless, user-documented diet data availability for analytical purposes tends to be constrained by lack of harmonization in collection of data and terms of use and privacy conditions (Maringer et al., 2018).

In case of developing nations, nutrition apps are mainly used at urban locations, and this may lead to a ‘digital divide’ if diet datasets from rural areas remain limited (Samoggia et al., 2021). Such a divide could be less pronounced in high-income countries. A novel digital tool that is being prototyped by CIMMYT (2022) in Bangladesh may address the possible rural-urban digital nutrition data divide in lower-income countries. This mobile app can be used by extension officers to assess diet data of smallholders to detect potential macro- and micronutrient deficiencies and suggest possible seasonal crops that can be grown in their homesteads to address the potential nutritional deficiencies in diets.

When used in larger scale, similar digital tools can also generate large-scale anonymised diet data that can be used for analytical purposes including modelling. The tool mentioned above helps extension officers to digitalize 7-day diet diaries recorded by farm household members. Digital tools that can lead to healthy food consumption while passively collecting food consumption data can prove useful for large-scale collection of diet datasets. Such datasets can be used for more comprehensive monitoring of national-level nutritional deficiencies and intervention targeting (Buckland et al., 2020) and policy simulations. In addition, if digital tools are used to source information on dietary supplements or additional constituents in foods (other than those meeting basic nutritional needs) that have health impacts, the role of bioactive dietary components (Yasmeen et al., 2017) in nutritional outcomes can also be explored using such datasets (Barnett & Ferguson, 2017).

Retailers’ data on food purchases by consumers is another emerging source of nutrition-related data (Saarijärvi et al., 2016). Digital data generated at points of sale (POS) can be used for consumption pattern recognition and then mobilized to promote healthier consumption behaviour. Application such as ‘NutriSavings’, a healthy grocery shopping reward programme in the United States, attempted to influence consumers’ decision to purchase healthy foods such as fruits and vegetables while reducing unhealthy fats, sugars and sodium using POS data (Nierenberg et al., 2019). Applications of point-of-sale nutrition data therefore could be advantageous in understanding nutrition consumption behaviour in developed countries and urban areas of developing nations.

Social media analytics (SMA) is also an emerging field for dietary data collection, behaviour analytics and population health assessments (Stirling et al., 2021). SMA is currently in data preparation and exploration phase, and future developments in quantitative data generation and analytics is expected to contribute towards nutritional surveillance and triangulation of other nutritional datasets. SMA applications in opinion mining, sentiment and content analysis and predictive analytics related to nutrition and health are emerging (Stirling et al., 2021). Initial studies on SMA show its high potential to yield policy insights similar to large surveys (Shah et al., 2020).
11.6 The Way Forward

The ideal (for computational social science applications) scenario (Fig. 11.2) is that all kinds of diet and nutrition data sources, such as the harmonized diet data from multi-indicator and other nutrition-related surveys, high frequency data from telephone surveys, diet data from mobile applications aiming nutritional advise, diet data from image-based diet detection applications, consumption information from point-of-sale data, diet- and nutrition-related social media data getting aggregated to a single data platform. These kinds of large datasets can be used for machine or deep learning as well as for simulation and modelling studies (e.g. FARMSIM) or for more conventional statistical analysis. Such national nutritional data-sharing platform can also include spatial data related to food supply chains, food environment and external factors so that major drivers can be diagnosed for observed diet patterns and for predictive analytics. The creation of unified platform for interoperable digital data can facilitate analytics that could in turn result in more insightful policy advice. The challenge is in the developing agreements with private firms and public agencies who are data holders to ensure the data access to researchers and the privacy of users and respondents. There is a need of policy innovations that encourage the creation of centralized data-sharing platforms, especially on nutritional data that allow researchers to analyse and derive policy insights that can lead to achievement of sustainable development goals.

Several recent studies showed the viability of using satellite data and mobile operator call detail records (CDR) to predict poverty levels at higher frequency and

![Fig. 11.2 Idealized national data platform for digital diet data and nutritional information. In case of data flows, the thick arrow shows the existing data stream, the thinner arrows represent novel data streams and dotted arrows represent upcoming and future data streams](image-url)
spatial granularity (Pokhriyal & Jacques, 2017; Steele et al., 2017). There are also attempts to track poverty using e-commerce data (Wijaya et al., 2022) and mobile money transactions data (Engelmann et al., 2018). Once large-scale geographic location-specific diet and nutrition datasets are available, it may also be possible to make predictions regarding diet patterns and nutritional deficiencies using satellite, CDR data, mobile money or e-commerce transactions. Social media data can also complement such nutritional surveillance. This kind of near real-time monitoring of agri-food systems components and nutritional analytics could potentially result in important new insights for policy, given the current infrequent and inadequate food consumption and nutrition-related data and limited capacity of the modelling tools.

The chapter provides an overview of significant challenges in agriculture and nutrition policy development and presents emerging digital data sources and computational social science methods that can potentially address the stated challenges. Given the right analytical framework, data platforms and enabling conditions, computational social science techniques can unravel the complex impact pathways to nutritional outcomes and contribute significantly to addressing the global burden of overweight, obesity, malnutrition, hunger and nutrition-related non-communicable diseases.

References


CIMMYT. (2022). *Nutrition smart homestead app*.


Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.