REQUEST FOR INFORMATION FROM THE GLOBAL DIETARY DATABASE, AUGUST 2019

1. OBJECTIVES OF THE GLOBAL DIETARY DATABASE (GDD)

- To estimate individual food and nutrient intakes worldwide by country, year, age across the lifespan, sex, education level, urban or rural residence, and pregnancy/nursing status.
- To use generated estimates to understand disease burdens, identify high-impact interventions, and evaluate and improve public health and nutrition policies.
- To create a public resource and dissemination platform for sharing intake estimates with all stakeholders in the global nutrition community.

2. LIST OF SURVEYS CURRENTLY INCLUDED IN GDD

Data collection for the Global Dietary Database has spanned roughly a decade. The first iteration of GDD, termed GDD 2010, modeled global diets through the year 2010. The current iteration of the project, termed GDD 2015, models global diets through 2015, as relevant data and covariates—e.g., from published reports, private Corresponding Members, FAO, and DHS (Demographic Health Survey)—are presently only available up to 2015, and estimating dietary intakes beyond 2015 would be wholly imputed and require strong assumptions about continuing trends.

As of June 2019, we have identified and retrieved 1,571 eligible survey-years of data from public and private sources. Of these, 1,216 have been checked, standardized, and approved for GDD 2015 model inclusion. The list of these surveys—including data on country, time, representativeness, assessment method, and sample demographics for each—is available in Appendix A. An additional breakdown of survey count by region is available in Table 1, below.

If nationally-representative surveys were not available for a country, we also considered national surveys without representative sampling, followed by regional, urban or rural surveys, and finally large local cohorts, provided that selection and measurement bias were not apparent limitations. [1] Surveys were included in the initial screening phase if the survey was reasonably population based and representative, exposure data were reported or could be plausibly obtained, and sample size included at least 100 individuals. For countries with no surveys identified, other sources of potential data were considered, including the WHO infobase, the STEP database, and household budget survey data. [1]

Most identified data were either privately-held or not in a format appropriate for our modeling. We thus relied almost entirely on direct author contacts in each country to provide us with exposure data directly. Where available, we accessed and downloaded public datasets, from which dietary exposure data was extracted, standardized, reviewed, and cleaned by GDD team members. [1]

Table 1. Characteristics of dietary surveys identified and retrieved for the 2015 Global Dietary Database, globally and by region.*

	Total (Global)	High-Income Countries	Sub-Saharan Africa	Middle East & North Africa	Asia	Former Soviet Union	Latin America & Caribbean	South Asia	N/A*
Survey characteristics, N									
Total survey-years of dietary intake retrieved	1571	428	290	132	225	208	180	83	25
Survey-years assessed and included	1216	376	210	95	170	158	127	56	24
Number of countries with data	185‡	24 of 24	44 of 48	20 of 20	24 of 24	28 of 29	29 of 32	7 of 8	12
Mean number of GDD dietary factors per survey	8	7	6	11	9	7	9	10	7
Assessment method, %									
Single or multiple diet recalls/records	22.9	32.2	6.7	18.9	38.8	15.8	12.6	28.6	12.5
Food frequency questionnaire	41.7	33.5	31.0	65.3	38.2	54.4	46.5	42.9	83.3
DHS questionnaire	15.8	0.0	48.1	13.7	8.2	6.3	31.5	25.0	0.0
Biomarker	13.2	21.8	14.3	2.1	14.7	4.4	8.7	3.6	4.2
Household Availability/Budget Survey†	6.4	12.5	0.0	0.0	0.0	19.0	0.8	0.0	0.0
Age group representation, %									
0-23 months	30.1	20.5	54.3	16.8	32.4	25.3	36.2	32.1	0.0
2-5 years	32.4	23.6	55.2	24.2	32.9	27.2	37.0	35.7	0.0
6-19 years	65.3	63.8	56.2	72.6	55.9	88.6	61.4	66.1	70.8
20-44 years	62.6	64.3	65.2	56.8	75.3	58.2	48.0	67.9	37.5
45-69 years	62.2	64.9	65.7	55.8	75.9	58.2	45.7	58.9	37.5
70+ years	32.6	38.6	21.0	33.7	37.6	42.4	20.5	30.4	8.3

*High Income Countries: Australia, Austria, Belgium, Canada, Cyprus, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Malta, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States.

Sub-Saharan Africa: Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cabo Verde, Central African Republic, Chad, Comoros, Democratic Republic of the Congo, Republic of Congo, Côte d'Ivoire, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, The Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, South Sudan, Namibia, Niger, Nigeria, Rwanda, São Tomé and Príncipe, Senegal, Seychelles, Sierra Leone, South Africa, Sudan, Swaziland, Togo, Uganda, Tanzania, Zambia, Zimbabwe.

Middle East & North Africa: Algeria, Bahrain, Egypt, Islamic Republic of Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Occupied Palestinian Territory, Oman, Qatar, Saudi Arabia, Syria, Tunisia, Turkey, United Arab Emirates, Yemen.

Asia: Brunei Darussalam, Cambodia, China, Fiji, Indonesia, Japan, Kiribati, Korea, Lao P.D.R., Malaysia, Marshall Islands, Micronesia, Myanmar, Papua New Guinea, Philippines, Samoa, Singapore, Solomon Islands, Taiwan Province of China, Thailand, Timor-Leste, Tonga, Vanuatu, Vietnam.

Former Soviet Union: Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Georgia, Hungary, Kazakhstan, Kyrgyz Republic, Latvia, Lithuania, FYR Macedonia, Mongolia, Montenegro, Moldova, Poland, Romania, Russia, Serbia, Slovak Republic, Slovenia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan.

Latin America & Caribbean: Antigua and Barbuda, Argentina, The Bahamas, Barbados, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, St. Lucia, St. Vincent and the Grenadines, Suriname, Trinidad and Tobago, Uruguay, Venezuela.

South Asia: Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka.

Countries yet to be assigned to a GDD region are labeled as N/A in the interim, including American Samoa, Anguilla, British Virgin Islands, Cayman Islands, Cook Islands, Greenland, Montserrat, Nauru, Niue, Puerto Rico, Saint Kitts and Nevis, Tokelau, and Tuvalu.

‡Global country count does not yet include countries unassigned to a GDD region.

†We recognize that numerous additional household availability/budget surveys are available, but these are significantly limited for assessing dietary intakes because they (a) assess purchasing rather than intakes or waste, (b) assess the whole household rather than individuals within the household whose intakes may differ substantially by age or sex, and (c) do not account for food purchased and consumed outside the home. We therefore only included household surveys when individual-level dietary survey data in a given country were not available for any year from 1980-2015. In such cases, we used the adult male equivalent (AME) method to account for age and sex when estimating intake per individual within a household.

3. DIETARY FACTORS COVERED IN DATABASE AND CODING METHOD

At its inception more than a decade ago, the GDD 2010 focused on dietary factors with confirmed or probable etiologic effects on major chronic diseases including cardiovascular diseases, diabetes, and cancer (22 dietary factors identified and included in total). GDD has since expanded in scope to characterize total diet in populations around the world, and now contains data on 55 dietary factors including 14 foods, 7 beverages, 15 macronutrients, and 19 micronutrients. We have also collected data on 2 indices of carbohydrate quality, but the very few sources will limit global modeling of these dietary factors. Identification and definition of an additional 4 dietary factors to be incorporated into future iterations of GDD have also been finalized. See Appendix B for the list of present and future dietary factors and their associated coding labels.

THE AVAILABILITY OF INDIVIDUAL FOOD VS. FOOD GROUP DATA

In the overall global model, the GDD determines and characterizes foods by group (e.g., fruits). Doing so maximizes the validity and accuracy of our prediction model and better facilitates comparisons of diet across countries.

Individual food intake data are also available in the GDD, including:

1. Original survey microdata:
    - Through correspondence with our private data owners, we have established data sharing agreements to publicly host their individual-level, non-modeled "microdata" on the GDD site. These microdata often contain dietary intake data for individual foods, which can vary widely between surveys. See part 8 for more information on our non-modeled microdata.
2. Newly-standardized microdata using FoodEx2:
    - See part 4 for information regarding GDD efforts, coordinated and harmonized with GIFT, to standardize and disseminate intake data for individual foods across surveys.

4. CODING METHOD USED FOR FOOD CATEGORIZATION

CATEGORIZING FOODS AS GDD VARIABLES

All dietary factors in the GDD have "optimal" and "suboptimal" definitions to maximize comparability of dietary intake across surveys and countries. Each individual survey undergoes a rigorous coding protocol to ensure all primary data are captured and incorporated into GDD as accurately as possible. The dictionary of optimal and suboptimal definitions for GDD dietary factors is available in Appendix C.

FOODEX2 CODING SCHEME

A major challenge in assessing dietary intake is the variation of descriptions of individual, self-reported food items, which can ultimately lead to assessment errors. To address this issue, we have developed methods and partnerships to apply FoodEx2—a sophisticated food description and classification system developed by the European Food Safety Authority (EFSA)—to surveys within GDD. This work will help to standardize global dietary intake beyond the dietary variables currently collected by GDD.

The GDD has signed MOUs and is formally collaborating with EFSA and FAO/WHO GIFT, as well as relevant data owners, in our work on FoodEx2 coding. The adaptation of the FoodEx2 system is a major advancement for the collection and storage of individual-level dietary intake data. With tools like FoodEx2, GDD more precisely informs global nutrition intakes, disease burdens, interventions, and policies.

5. INDICATORS OF ACTUAL (NON-MODELLED) DATA THAT CAN BE GENERATED FROM THE DATABASE

There are currently no plans for GDD to generate separate indicators from non-modeled data. However, as both non-modeled data and FoodEx2 standardized data are approved for hosting on the GDD site by the original owner, GDD site users will be able to download these data and construct indicators themselves. See part 8 for more information about the approval process for hosting non-modeled data on the GDD site.

6. EXACT SPECIFICATIONS OF MODELS USED TO GENERATE ESTIMATES OF INDIVIDUAL INTAKE OF FOODS OR NUTRIENTS AT NATIONAL LEVEL

The GDD prediction model estimates mean intakes of 55 dietary factors, jointly stratified by country, year, age, sex, urbanicity, education, and pregnancy status, in 185 countries by synthesizing survey mean intake data and relevant covariates from a range sources. The Bayesian multilevel framework has some advantageous properties that are appealing for our aims. A summary of the model is provided in Appendix D.

The prediction model is further informed by a broad range of covariate data from various global sources. These include:

- FAO food balance sheets data, 1980-2013
- Principal component analysis of FAO and GENuS data, 1980-2013
- World Bank GDP, 1980-2015
- Unemployment rate, 1991-2015
- Education years, 1980-2010
- Gini coefficient, 1980-2015
- Poverty rate, 1991-2015
- Harvard Global Expanded Nutrient Supply (GENuS) data, 1980-2013
- Industry sales data on fat consumption, 1998-2015
- Precipitation, 1982-2014
- Land area
- Latitude
- Coastline ratio

7. INDICATORS OR TYPES OF MODELLED ESTIMATES THAT CAN BE GENERATED FROM THE DATABASE

GDD 2015 estimates mean levels of dietary intake for 55 dietary factors across the global population, jointly stratified by the following characteristics:

1. Year
    - 1990 through 2015

2. Country
    - 185 countries
3. Age groups
    - 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+ years
4. Sex
    - Male and female
5. Residence
    - Urban and rural (as defined by each survey's characteristics)
6. Education level
    - Low (0-6 years formal education), medium (>6 to 12 years), and high (>12+ years)
7. Pregnancy status
    - Currently nursing, currently pregnant

GDD modeled estimates can also be used to generate different dietary patterns and metrics, such as healthy and unhealthy diet global patterns as well as other common metrics (with some modifications) such as the Healthy Eating Index (HEI), Alternative Healthy Eating Index (AHEI), Mediterranean Diet Score (MDS), Dietary Approaches to Stop Hypertension score (DASH), Infant and Young Child Minimum Dietary Diversity (IYCMDD), and Minimum Dietary Diversity for Women (MDDW).

8. HOW TO ACCESS THE RAW DATA IN THE DATABASE

A key aim of the GDD is to become an archive of all the surveys which serve as our model inputs. Such a functionality will be helpful in connecting researchers with one another and with important data on the populations they study. This functionality relies on significant collaboration from the original data owners. We have collected and continue to actively pursue signed data sharing agreements (DSAs) with survey owners to ensure the safe and responsible sharing of their raw data (microdata). We anticipate this process will be completed as scheduled by June 2020.

Appendix A contains information regarding current sharing status. As of November 2019, the GDD model includes:

- 871 publicly-available surveys
- 345 privately-held surveys
    - 98 with signed DSA agreements to allow public download
    - 1 available upon further request and approval by the data owner
    - 246 not yet available
        - 215 with ongoing communications for a potential DSA
        - 31 declined to share publicly

GDD site users will be able to access general survey information about all surveys used for GDD input data, with or without DSAs, via an interactive global data availability map. This map is currently being constructed and will be updated on a rolling basis to reflect updates in DSA status. Users will be able to

filter this map display by survey characteristics (assessment method, subject age range, representativeness, etc.) to isolate countries and surveys of interest. Surveys for which a DSA is in place will have an associated, freely-available dataset which users can download or request (when applicable). For publicly available datasets, links to the site from which the dataset originates will be listed in lieu of data files, which generally cannot be directly shared based on the original data sharing requirements of the owners.

9. PLANS FOR INCLUSION OF ADDITIONAL SURVEYS TO YOUR DATABASE:

To allow completion of the findings, the incorporation of surveys into the GDD 2015 modeling phase has closed. Any additional surveys which may be identified will be reserved for a future round of funding and efforts (e.g., GDD 2020), unless they are so large or important to warrant inclusion now. In addition, as household-level surveys are less reliable than individual-level ones, we have deferred adding any further surveys to the current model (some household-level surveys are currently from large countries, such as Russia, for which individual-level data are not available). See Appendix E for the full list of surveys we have identified but not yet analyzed.

10. ANY VALIDATION WORK THAT HAS BEEN DONE ON:

ACTUAL VS. MODELLED INTAKES, FOR DIFFERENT POPULATION GROUPS

For GDD 2010—the project's most recent iteration—model validity across different iterations was evaluated using cross validation, randomly omitting 10% of the raw data and comparing the imputed intakes with the original raw data. [2]

For GDD 2015, we are currently completing work for 5-fold model cross-validation across all dietary factors to formally test the predictive ability of the GDD 2015 prediction model. In this established model validation method, the model raw data inputs are randomly divided into five groups ("folds"). In each round of testing, one group is dropped, the model is imputed based on the remaining data, and the ensuing predictions are then compared to the omitted raw data. This process is then repeated four more times for the remaining four groups, and an overall statistical estimate of model fit is generated. A summary of this work will be provided upon its completion in 2020.

DIETARY PATTERNS/FOOD GROUP INTAKES VS. ADEQUACY OF NUTRIENT INTAKE

Currently, the GDD prediction model is designed to estimate mean dietary intake across subgroups by country-year, synthesizing information from various types of surveys. We also plan to expand our work to estimate the usual intake distribution—that is, to estimate usual dietary intake from just the means and standard deviations of intake provided to us by contributing members.

The general philosophy of our approach is to account for systematic bias from assessment methods of surveys used in our prediction model to obtain the full usual dietary intake distributions. By creating and fitting a separate mean-to-standard deviation linear regression model on the dietary data, we can

predict standard deviations from mean estimates for each strata and country-year, thereby generating joint predictive distributions of intake.