Harnessing the power of the census: characterising wastewater treatment plant catchment populations for wastewater-based epidemiology

Benjamin J. Tschärke^{1*}, Jake W. O'Brien, Christoph Ort², Sharon Grant¹, Cobus Gerber³, Richard Bade³, Phong K. Thai¹, Kevin V. Thomas¹, Jochen F. Mueller¹

¹ Queensland Alliance for Environmental Health Science, The University of Queensland, 20 Cornwall Street Woolloongabba, Queensland 4102, Australia

² Eawag, Swiss Federal Institute of Aquatic Science and Technology, CH 8600, Dübendorf, Switzerland

³ School of Pharmacy and Medical Sciences, University of South Australia, Adelaide 5001, Australia

***Corresponding Author:** Email address: b.tschärke@uq.edu.au (Benjamin Tschärke)

Abstract:

Wastewater studies that provide per-capita estimates of consumption (influent) or release (effluent) via wastewater systems rely heavily on accurate population data. This study evaluated the accuracy of Wastewater Treatment Plant (WWTP) reported populations, as well as hydrochemical parameters, against accurate populations from a population census. 104 catchment maps were received from WWTPs, geo-located in geospatial software and overlaid with the smallest area unit of the Australian census, equating to 14.9 million Australians or 64% of the national population. We characterised each catchment for population counts, as well as by age profile, income profile and education level. For a subset of sites, population estimates using hydrochemical parameters BOD, COD and dissolved ammonia were evaluated for accuracy against census populations. Compared to census-based estimates, population estimates provided by WWTP personnel were overestimated by 18% on average. Similarly, hydrochemical-based population estimates had high RSD (> 44%) for BOD, COD and ammonium between sites, suggesting their applicability for use in population estimation may not be appropriate for every WWTP. Surprisingly, 46% of catchments had a skewed age distribution; 6% skewed older and 40% younger. Through this process WWTP catchment populations can be characterised in a way which will enhance the interpretations of per-capita estimates.

This document is the accepted manuscript version of the following article:

Tschärke, B. J., O'Brien, J. W., Ort, C., Grant, S., Gerber, C., Bade, R., ... Mueller, J. F. (2019). Harnessing the power of the census: characterizing wastewater treatment plant catchment populations for wastewater-based epidemiology. *Environmental Science and Technology*, 53(17), 10303–10311. <https://doi.org/10.1021/acs.est.9b03447>

Keywords

Population estimation; wastewater catchment; wastewater analysis; population characterisation; census; demographics

Highlights:

- 104 WWTP catchment populations were estimated from Census data and catchment maps
- Hydrochemical population estimates had high RSD > 44% for BOD, COD and ammonium
- WWTP's overestimated service population by 17.7% in this study
- Population estimates had *de jure* population uncertainty of 1.8% on average
- 46% of catchments had a skewed age distribution

Graphical Abstract

1. Introduction:

Wastewater-based epidemiology (WBE) is a rapidly evolving technique, gaining significant public attention and acceptance for its near-real-time population monitoring of drugs of concern such as illicit drugs and pharmaceuticals as well as exposure to chemical and biological agents¹⁻⁴. Estimating the population contributing to wastewater samples is one of the most critical parameters used in the calculation of per-capita normalised estimates⁵⁻⁷. For example, wastewater drug loads normalised to per-capita consumption estimates enhances the utility of the data by enabling the direct comparison of findings among wastewater catchments, both locally and internationally. This is particularly important considering the increasing number of WBE studies conducted and the extensive use of this per-capita consumption data by associated fields tangential to analytical chemistry. Despite the vital nature of the population on these final per-capita estimates, the accuracy is often difficult to estimate as the specific contributing population can be difficult to quantify.

The population size of wastewater treatment plant (WWTP) catchments can be estimated in several ways. Typically, an estimate of the residential population living within the boundary of a wastewater catchment is provided by the wastewater treatment authority. These populations are generally estimated by either the census population for the whole city (not necessarily specific to the catchment), WWTP masterplan projections, the number of water connections or the total flow volume received at the treatment plant. The population can also be estimated analytically using hydrochemical data which can be measured onsite such as nitrogen, phosphorous, Chemical Oxygen Demand (COD), biological oxygen demand (BOD) and dissolved ammonia from urea in urine⁸. While these methods are quite common, the per capita loads and volume of wastewater generated per person of these remain uncalibrated. Nitrogen, phosphorous, BOD and COD loads may also be influenced by industrial input, and

flow volumes may be impacted by storm water runoff or via sewer infiltration or exfiltration. To combat these issues, exogenous biomarkers such as acesulfame, caffeine and pharmaceuticals⁹ or nicotine metabolites can be co-analysed to evaluate short term changes in population¹⁰. In addition, endogenous biomarkers such as the neurotransmitter metabolite 5-hydroxyindoleacetic acid (5-HIAA), cholesterol, coprostanol, or creatinine¹¹, have also been explored for their applicability, with mixed results. For example, in Lai 2011⁷, population uncertainty was evaluated based on prescription pharmaceuticals; derived from pharmaceutical mass loads and national statistics on pharmaceutical consumption. However, this assumes the uniform distribution and consumption pattern of pharmaceuticals across different sites, which might not hold true for pharmaceuticals with age bias such as atenolol. This may be problematic for sites that are heavily skewed for age.

The drawbacks of many of these approaches is that the use of lifestyle products or pharmaceuticals are culture dependent or could be affected by availability and marketing. Within countries, different prescribing practices or preferences might exist. Even where the exogenous markers have been calibrated for a specific catchment⁹, levels of some proposed population biomarkers are significantly variable or may fluctuate seasonally (e.g. acesulfame)¹². Furthermore, knowledge of the population and population demographics contributing to a wastewater sample is an influential factor for the interpretation of WBE results, and this is simply not possible to obtain via these approaches.

Novel population estimation methods include mobile phone triangulation based on anonymised mobile phone pings⁶. Such telephonic methods are valuable to account for short term changes in population and can reveal some demographic information such as gender and age, yet can be prohibitively expensive (this costs approx. AUD \$5,000 for 1 week at

hourly resolution for part of one city). Currently the technology can be limited as it was not yet available to isolate data to specific regions such as the discrete area of a WWTP catchment in Australia.

Herein we describe a methodology to estimate and characterise the residential catchment population of WWTP catchments by overlaying WWTP catchment maps against publicly available geo-referenced census population data in geospatial software. This has additional benefits over hydrochemical-based, population-biomarker, and telephonic phone ping estimates, as catchments can be evaluated in detail for demographics such as age, income and education level specific to the catchment area. In addition, population counts can be estimated with a greater resolution and accuracy than those generally provided by WWTPs. If census is unavailable, layering other types of datasets against geo-referenced catchment maps allows for enhanced comparisons. For example, location-based policing or health data can be collated and compared specific to the catchment area which can give enhanced context to WBE results and assist in interpretation.

This study aimed to evaluate the following; 1) to describe a WBE methodology to estimate and characterise catchment populations using census; 2) Determine the accuracy of WWTP-provided population estimates; 3) Determine the accuracy of hydrochemical-based population estimates; and 4) Explore how census information can be used to characterise WWTP catchment populations by quantifying population change, age profile, income profile and education level. These aims were achieved by estimating and characterising the population of 104 WWTP catchments in Australia and comparing these to WWTP-provided estimates as well as hydrochemical-based estimates. To demonstrate an application of the

113 methodology, high-resolution age-distribution profiles were calculated for each catchment
114 and evaluated for age skewness, income and level of education profiles.

2. Methods:

2.1. Wastewater catchment information

Wastewater catchment information was collected during a wastewater sampling campaign organised to coincide with the 2016 Australian Census, as described by O'Brien et al.¹³. Data from wastewater sampling questionnaires were acquired for 110 WWTPs across Australia, including information on the WWTP processes, hydrochemical concentrations, the WWTP estimate of the population served, and a sewer catchment boundary map (see O'Brien et al. SI for details). As a requirement of data use, the specific locations of the WWTPs are kept confidential for this study, and referred to by site codes.

Population estimates were gathered from participating WWTPs via the WWTP processes and catchment questionnaires. These questionnaires asked for the estimated number of inhabitants within the catchment boundaries as well as the methodology used to derive the estimate. The methodologies used to derive the population estimate were consolidated into 7 categories: Not Reported, WWTP census or council population estimate, WWTP estimate (not specified), number of water connections, flow volume, projection or planning or forecasting estimate, or number of privately-owned wheeled rubbish bins, termed 'wheelie bins' in Australia. As wheeled rubbish bins are allocated to each premise in these areas, the number of bins is multiplied by the average household population, between 2.5 and 3 people per household for those sites. To evaluate the accuracy of each population estimate method, these categories were then compared to the final population estimate from this study, described in Section 2.2.

2.2. Population estimation of each WWTP based on census

High resolution catchment maps were requested through the questionnaire from the WWTPs and were received from 104 WWTPs either as Geographic Information Systems (GIS) files, PDF files, or image files, albeit of various quality. Some catchment maps required georeferencing into GIS software, using QGIS version 2.18.10. Specifically, the catchment boundary maps were geographically assigned to a location by the identification of landmarks such as cross roads and intersections across many points within the provided map, against a geo-located map of Australia (projection GDA-94). This linked the points in the provided file to the same points in the geospatial software. Maps were visually verified for fit by comparing landmarks in the catchment map to the same landmarks on the satellite imaging layer within QGIS. Georeferenced maps were then digitised into catchment boundary map polygons to generate one layer with all 104 catchments. An example catchment is shown in Figure S1. It was not possible to verify the accuracy of maps provided by the WWTPs. Further detail on the population estimate methodology using catchment maps and census data is available in the supplementary information.

2.3. *De jure* population uncertainty and accuracy estimates

To compare the statistical relevance of findings in sections 2.3.1, 2.3.2, 2.4, and 2.6, an ANOVA with Tukeys post-hoc comparison test was conducted on the log-transformed data to determine significance among WWTP metadata categories.

2.3.1. *De jure* population uncertainty estimate

The total uncertainty of the *de jure* population based on census (excluding population movement) and correlation with catchment boundary maps is due to the following

uncertainties: 1) uncertainty of the census population counts due to survey deployment; and
2) uncertainty associated with the catchment boundary, particularly for the population
distribution at the margins.

As the Australian census has a very high response rate (>95%)¹⁴ and low relative standard
error (0.9%)¹⁵ the uncertainty of the population survey is considered negligible. More
information is provided in the supplementary materials.

For uncertainty relating to 2), there is negligible uncertainty for the census areas fully
contained within the catchment boundary (0.9% RSE of the census counts). However, for the
areas which intersect the catchment boundary, the effect of using the area ratio to calculate
the proportional population can be evaluated by determining the maximum and minimum
expected populations for the catchment. Expressing the population of the uncertain areas as
a percentage of the total population estimate gives an indication of the uncertainty of the
total *de jure* population. These concepts are described in more detail in the supplementary
materials.

2.3.2. Accuracy estimate

Population estimates provided by the WWTPs were compared with the high-resolution
census estimates calculated in this study (section 2.2). Results were expressed as the
percentage of the census estimate. The accuracy estimates were grouped into the same
categories as section 2.3.1 to evaluate accuracy based on WWTP metadata.

2.4. Population estimate using Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), ammonium ions and flow.

Wastewater parameters of BOD, COD and ammonium ions in mg/L were determined by participating WWTP laboratories and reported in the questionnaire for up to 7 days during the week of census in 2016 (for a detailed example of the deployed questionnaire, see ¹³). A total of 17 sites reported this information for BOD, 18 for COD and 22 for ammonium. Although flow data was collected for 83 sites for up to 7 days, for consistency the set of the 24 sites that reported either BOD, COD or ammonium data was also selected for the estimate based on flow. Population estimates were generated from ammonium parameters following a published methodology ⁸. For the BOD and COD estimate, per capita equivalents were calculated using 54 g/d/person for BOD and 120 g/d/person for COD ⁸. Ammonium ions were calculated using both the lower bound of 7.7 g/d/person and the upper bound of 8.5 g/d/person as reported in Been et al ⁸. Based on the census population calculated in this study, the average per-capita ammonium ion loads were also calculated. A flow contribution of 250 L/day/person was also assessed for accuracy ⁹. All parameters were expressed as a percentage of the census population.

As some of the WWTPs used flow to determine their WWTP catchment population estimate, the average daily per-capita flow was calculated from 7 days of flow data and the newly calculated population estimates on the week of census 2016. The per-capita flow was also contrasted against WWTP population size category and greater capital city area or regional area location to determine their influence on accuracy. All WWTP catchments assessed for flow were sewer systems separate from surface water.

2.5. *De jure* population change of the catchments between 2011 and 2016 census

Population change was estimated by contrasting the catchment maps against the 2011 and 2016 census mesh block population counts to determine if the population changed across the 5 years (as per section 2.2). Annual population growth/decline was expressed as the annual percentage change from 2011 estimates assuming linear change. Changes in catchment boundaries between these years were not accounted for in this assessment as backdated 2011 catchment maps were not available.

2.6. Population movement estimate – *de jure* and *de facto* populations on census day.

To estimate the potential for population movement for each catchment, the *de jure* (usual resident population) and *de facto* (present on census day) populations were collected per mesh block from the ABS census 2016 and calculated in the same manner as per Section 2.2. The percentage difference between these estimates was calculated and used as a measure of potential population movement. This data will show if there was a net difference of expected (*de jure*) vs actual (*de facto*) population within the catchment.

2.7. Determination of population age, education and income profiles

The age, education and income profiles of each catchment were calculated from ABS data, explained in detail in the supplementary information. Results were expressed as the percentage of the total at each category, to enable comparison between catchments, and to prevent site identification. The age population distributions were assessed for Pearson's second coefficient of skewness to determine if a catchment's population distribution was skewed to either a younger or older age. This was achieved using skew thresholds of 0.367 and -0.367 which define the 90% confidence interval for sample sizes of $n=100$ ¹⁶.

226 Distributions were plotted using R statistical software (version 3.5.2) along with the general
227 additive model (gam, formula = $y \sim s(x)$) smoothed line of best fit. The WWTP catchment codes
228 used in this study correspond to those used for the National Wastewater Drug Monitoring
229 Program¹⁷ for the sites included in both studies.

3. Results and discussion:

3.1. Questionnaire return and responses

In addition to the 104 catchment maps received, 99 WWTPs provided population estimates, 82 provided daily flow rates, 62 made attempts to complete the WWTP processes and catchment questionnaires, 17 sites provided BOD concentrations, 18 sites provided COD concentrations, and 22 sites provided ammonium concentrations. These data were used for further calculations.

3.2. WWTP catchment population based on census

The catchment populations estimated for the 104 WWTPs ranged from small to large populations, Figure S1, with the largest proportion of sites in the 20,000 to 50,000 inhabitants' range. The distribution across categories was relatively even within each category, with each containing at least 5% of sites. The total population coverage was 14.9 million, which is 64% of the Australian population (Census population of 23,401,892 in 2016)

3.3. *De jure* population uncertainty and accuracy estimates

3.3.1. *De jure* population uncertainty estimate

The WWTP catchments in this study had an average *de jure* population uncertainty estimate of 1.8% for catchments that provided catchment maps (for a full table see Table S1 in supplementary Information). There were some differences observed between categories of WWTP location and size, Figure S1. However, these differences were not statistically significant between capital and regional categories (ANOVA with Tukeys post hoc test, $P > 0.05$). This may not be expected as census units in regional areas are larger due to the lower population density. *De jure* population uncertainty using best case scenario (Australian

example) can be reduced by utilising the smallest resolution of a National Census and WWTP catchment boundary maps. The uncertainty was on average 1.8% of the total population (range 0.1% - 15.4%).

3.3.2. Accuracy of the WWTP estimates

To determine the accuracy of the WWTP catchment estimates provided by WWTP authorities, the population estimates were expressed as a percentage of the census estimate from this study and compared by WWTP metadata, Figure 1. The accuracy of WWTP estimates ranged from 32% to 268% and was not significantly influenced by catchment population category or location (capital or regional). Methodology for WWTP catchment population estimation did show some variation. The municipal “wheelie bin” data and influent flow were among the most accurate for estimating the *de jure* population, with BOD or COD load-derived population estimates (determined by the WWTPs) seemingly least accurate. However, no significant difference was determined between any category (ANOVA with Tukeys post hoc test, $P > 0.05$). Two data points were excluded due to very high differences, 551% and 448% difference in population accuracy, which was likely due to population movement (tourism, movement of population of 30% and 40% of the total residential population, respectively), described further in section 3.6. It is important to note the purpose of the WWTP population estimates are for operational use from a treatment perspective. Therefore, the intention of the estimates provided by the WWTPs may not be to predict the service population, but to serve as a guide for the pollution loads or water volume expected during treatment.

Figure 1: Accuracy of the WWTP-provided estimates as compared to the census-based estimates. a) Accuracy is categorised by the population size of the WWTP catchment (in graph order $n = 7, 9, 9, 16, 27, 11, 8, 9, 4, 4$), b) capital or regional location ($n = 32, 72$), and c) by the WWTP population estimate method ($n = 7, 4, 10, 11, 42, 27, 2$). The average WWTP population estimate was 17.7% higher than the census estimates of this study (red dotted line, 117% accuracy).

This study suggests that the average accuracy of WWTP estimates are within 18% of the census value. However due to a large variation in accuracy between sites (range 0.2% - 95.5% difference in population estimate), this should be calculated on a site-by-site basis using the best available methodology. Although there were no observed differences between means for many of the methods, based on the IQR and range, a suggested order of data quality for other studies may be: census estimate > flow > forecasting estimate > number of connections > BOD/COD. Due to the low accuracy, a population calculated using BOD/COD may require verification via other estimation methodologies. Quarterly and/or annual updates for census data may be useful, where available, for normalising and extrapolating population estimates over time.

3.4. Accuracy of population using water quality measurements

In order to determine the accuracy of hydrochemical-based population estimates, these were expressed as a percentage of census estimates from this study, Figure 2. Population estimates using BOD, COD, ammonium ions and flow rate were evaluated, and correlated against census population from this study. No significant difference was determined between any measure (ANOVA with Tukeys multiple comparison post hoc test, $P > 0.05$). Although the mean value for BOD, COD and flow was reasonably accurate, a large variance was observed within each measure. This equated to relative standard deviation (RSD%) values of; 49%, 44%, 44%, for BOD, COD and Flow, respectively. However, variation for ammonium-based population estimates was higher and less accurate on average, with RSD% values of 55% for ammonium (RSD values calculated with outliers removed).

Figure 2: a): accuracy of population estimates using hydrochemical measures versus census-derived values. Hydrochemical parameters were of the subset of catchments that provided b) ammonium data c) BOD loads, and d) COD loads. Ammonium nitrogen has been estimated using the ^ low (7.7g/d/p) and # high (8.5g/d/p) excretion range estimated in Been et al⁸, and * using the average daily per-capita excretion of the catchments measured in this study (average of 9.24g/d/p). For clarity, outliers of 404 & 314, 366 & 285, 337 & 262, were removed from graph 2a for low, medium and average NH4+ respectively.

The per capita wastewater flow was calculated for each of the metadata categories, Figure 3.

The average per capita flow of catchments that reported flow (n=83) was 293.3 L/day/person, with an 95% upper bound of 550 L/day/person and a 95% lower bound of 150 L/day/person, with an overall RSD of 43%. One outlier was removed at 974 L/day/person. This catchment is the only sewer catchment in this study known to have water influx from a river system and was excluded for this reason.

316 Highest variation in per-capita flow appeared to be among the 10,000 – 50,000 population
317 catchments, with lower variance in population categories above 50,000, Figure 3a. These
318 results show that using hydrochemical measures as a comparator between studies and
319 locations, even within the same country, state or city, may be potentially limited due to
320 significant differences observed in per capita water use among catchments. However, they
321 may be comparable to WWTP provided estimates in some instances, albeit with a larger
322 variance. It may be necessary to calibrate BOD/COD/ammonium loads on a site-by-site basis
323 if within-country data is unavailable. Large variations in per-capita flow show that dilution can
324 vary up to 10-fold between WWTP catchments (~90 - ~900 Litres per person per day), which
325 means that influent concentrations may not be a good comparative measure between
326 locations. Normalising to flow by converting to loads is recommended to account for
327 differences in water use.

Figure 3: The daily per capita consumption of water (L) within each catchment, categorised by a) catchment population size, b) capital or regional designation, and by c) WWTP population estimate. d) Population vs flow (left) and e) per capita flow of each regional and capital city catchment, categorised by population size.

3.5. De jure population change of the catchments between 2011 and 2016 census

For temporal WBE studies over several years, long-term changes in population have the potential to skew results. To evaluate long term changes in the WWTP populations, the 2011 and 2016 *de jure* and *de facto* census populations were estimated and used to determine the

monthly and the average yearly increase between the two censuses. For each census the *de jure* and *de facto* populations were captured by the Australian census questionnaire.

Annual population change showed ~75% of all sites increased by a relatively small margin (0-6% over 5 years) but this appeared more variable in regional areas, with larger increases on average in capital areas, Figure S4. More information is also listed in the supplementary information. For temporal WBE studies it would be beneficial to check for significant population change in each catchment (only 3% of catchments grew by >6%, yet this may not be the case elsewhere). Differences in *de facto* and *de jure* populations suggest population movement influences the population estimate mainly in smaller populations in tourism-designated areas (all catchments with greater than 10% population difference), with negligible change expected for the remaining catchments (less than 5% difference). National and international tourism may need to be investigated for influence on population-normalised mass loads, particularly for smaller catchments, or for season. A *de facto* population model could be used in these instances to estimate daily population based on relative change from the census calibration (e.g. a pharmaceutical-based population model, mobile phone pings or site-based BOD or COD calibration).

3.6. Potential population movement: *de jure* and *de facto* populations on census day.

The percentage difference between *de jure* and *de facto* populations was highest for catchments smaller than 50,000 population, and slightly higher in regional areas compared to capitals, Figure S5. The clear majority of catchments had little difference between *de jure* and *de facto* populations (83 catchments had less than a 5% difference, 7 catchments had 5-10%, 9 catchments had 10-20% and 5 catchments had 20%-40% difference in population), with an average of 4.4% difference (max = 40.5%, min = 0.03%). Population variations were mostly

observed for smaller catchments. Thirteen of the fourteen catchments that had more than 10% difference had populations below 50,000, while the remaining catchment had a population between 200,000 to 500,000 (population difference of 13%). All of the catchments with greater than 10% variation are considered tourism destinations ¹⁸. The top 3 treatment plants that were affected most by population movement (30-40%) used flow or BOD or COD load-derived population estimates. In areas with greater tourism, these measures may be better to predict pollution loads or required service capacity. Significant differences ($P = 0.0395$, ANOVA with Tukeys posthoc comparison test) were observed between 5,000 – 10,000 and 100,000 – 200,000 population categories and for capital vs regional ($P=0.0115$, two tailed students t test), showing smaller, regional populations were influenced to a greater extent by population movement.

3.7. Potential application of geo-referenced WWTP catchment maps

As an example of the application of the census-based methodology, the population within WWTP catchments were characterised by their age distributions in one-year age groups. No major population skew was found for 54% of catchments, while 40% of catchments had a positive skew toward younger populations and 6% of catchments had a negative skew toward older populations. Three examples of the skew types are shown in Figure 4. The population distribution of the catchments could be used to determine reasoning behind consumption rates of drugs and pharmaceuticals as many will be age-dependent. For example, the catchments shown at right in Figure 4 (A) had particularly high opioid consumption in other studies ¹⁷, which may, in part, be influenced by the older population distribution. Older age groups may be cohorts requiring a higher degree of treatment with analgesics. Similarly, education level or income profiles may also be useful in interpreting WBE consumption

estimates. For example, site 8 has a much higher income level and education level (Figure 4 (B,C)), and also has the highest cocaine consumption of WWTPs assessed in Australia ¹⁷. A combination of demographic factors such as age, income, education, employment status or socioeconomics will contribute to drug use ¹⁹. A detailed evaluation of these issues will be the subject of further publications.

The higher resolution will also allow for the accurate normalisation of drug consumption estimates to populations within a specific age range, which can assist in triangulating with other datasets that normalise results for specific age ranges. Such ranges could also be applied in a bespoke way for each drug. For example, pharmaceuticals that are treatments for specific diseases that appear later in life, such as mass loads of selegiline for Parkinson's disease (typically diagnosed for patients over 40 years old) could be normalised to age > 40 to account for observed skew between catchments. Alternatively, illicit drugs which are most likely consumed by population within age ranges 15-65, could also be normalised to the population between these values. This is particularly relevant if consumption estimates are skewed low due to an ageing population. Overall, one might expect wastewater treatment plant catchments to be non-skewed due to the large areas and populations. However, results here show that there can be significant population skews for catchments under study and this should be evaluated in WBE studies that compare catchments, particularly for analytes which may be age interdependent. Based on these observations, it may not be safe to assume WWTP catchments to be homogenous in age distribution, income profile or level of education profile. When comparing WBE consumption or exposure patterns between sites these may need to be taken into consideration as potential points of influence. The age, education and income level profiles for each catchment is provided in the supplementary material. All site

406 codes refer to WWTPs in the Australian Census sampling ¹³ and the Australian National
407 Wastewater Drug Monitoring Program studies ¹⁷.

408

409

410 Figure 4: A) Population distribution graphs showing the proportion (%) of the population at
 411 each age for three example catchments. Positive skew (left), no major skew (centre) and
 412 negative skew (right) are shown. No major skew was found for 54% of catchments, while
 413 40% of catchments had a positive skew (younger population) and 6% of populations had a
 414 negative skew (older population). B) Examples of catchments with lower education
 415 distributions (left) to higher education distributions (right). Where Cert = Certificate, Adv. =
 416 Advanced, Dip. = Diploma, and Grad. = Graduate. C) Examples of income distributions for
 417 lower income catchments (left) and higher income catchments (right) in Australian dollars
 418 (\$AUD).

419 **3.8. Relevance for other countries:**

420 The WWTP catchments assessed in this study account for 64% of the Australian population
421 (14.9 million people), and the distribution of WWTPs per population category showed this
422 study had a good representation of sites from several population categories (see Figure S2).
423 It is worth noting that the data presented here may not be representative of the situation in
424 other countries.

425 The main limitations to this methodology is the frequency of national censuses. The United
426 Nations statistical division recommends a census should be conducted every 10 years, with 5
427 year intervals for increased data quality ²⁰. The data collection frequency, quality and
428 availability vary from country to country. In the next couple of years, several countries have
429 a population census planned ²¹. For example; in 2020 The United States of America, Mexico,
430 Brazil, China, Indonesia, Malaysia, United Arab Emirates and Russia (among others) have
431 national censuses planned, while in 2021 Australia, South Africa, Canada, Greenland,
432 Bangladesh, China (Hong Kong), Turkey, Croatia, Estonia, Finland, Greece, Hungary, Iceland,
433 Italy, Portugal, Slovakia and Switzerland (among others) have censuses planned. This would
434 be a great opportunity for calibrating population biomarkers and/or population estimate
435 methodologies in those countries. A population estimate methodology (pharmaceutical
436 loads, ammonium, BOD, COD, Flow, mobile phone pings) can be calibrated during a
437 population census. For periods between censuses, the relative change in the other measures
438 could be used to estimate population at higher time-resolution.

439 A breakdown of the workflows utilised in this study and rough timelines is outlined in Figure
440 S6. For data processing this was completed by the authors in this study, with outside help to
441 learn the digitisation/geo-referencing process. However, for other studies it may be easier to

consult or collaborate with geospatial or geography departments at your university to process the census data. To obtain geolocated census data, this would be best to perform a web search or contact the geography department of your university, the National census bureau of your country, or the National statistics department, to determine what census units are available and the resolution. It is possible that census may be available to suburb, postcode or zipcode or other location units. The collation of these data are not limited to population censuses; the principles outlined in this study are applicable to any geospatial dataset which may contain or infer population or socio-demographic data.

4. Acknowledgements

We gratefully acknowledge the wastewater treatment authorities for providing catchment maps and responses to questionnaires. The authors wish to thank Andre Dunn-Johnstone for his assistance with QGIS software. The Queensland Alliance for Environmental Health Sciences, The University of Queensland, gratefully acknowledges the financial support of the Queensland Department of Health. This project was supported by an Australian Research Council Linkage Project (LP150100364).

5. Supporting Information

There are four supplementary documents. The first contains supplementary graphs, WWTP metadata and explanatory text, as mentioned in the manuscript. The second, third and fourth are PDFs contain the age distribution, income distribution and education level distribution figures of all sites, respectively.

6. References

1. Choi, P. M.; Tschärke, B. J.; Donner, E.; O'Brien, J. W.; Grant, S. C.; Kaserzon, S. L.; Mackie, R.; O'Malley, E.; Crosbie, N. D.; Thomas, K. V. Wastewater-based epidemiology biomarkers: Past, present and future. *Trends in Analytical Chemistry*. **2018**, *105*, 453-469.

2. van Nuijs, A. L.; Castiglioni, S.; Tarcomnicu, I.; Postigo, C.; de Alda, M. L.; Neels, H.; Zuccato, E.; Barcelo, D.; Covaci, A. Illicit drug consumption estimations derived from wastewater analysis: a critical review. *Science of the Total Environment*. **2011**, *409*, (19), 3564-3577.
3. Gracia-Lor, E.; Castiglioni, S.; Bade, R.; Been, F.; Castrignanò, E.; Covaci, A.; González-Mariño, I.; Hapeshi, E.; Kasprzyk-Hordern, B.; Kinyua, J. Measuring biomarkers in wastewater as a new source of epidemiological information: Current state and future perspectives. *Environment International*. **2017**, *99*, 131-150.
4. Castiglioni, S.; Bijlsma, L.; Covaci, A.; Emke, E.; Harman, C.; Hernández, F.; Kasprzyk-Hordern, B.; Ort, C.; van Nuijs, A. L.; de Voogt, P. EMCDDA Insights, Estimating community drug use through wastewater-based epidemiology. **2016**, *22*, 17-33. DOI 10.2810/017397
5. Castiglioni, S.; Bijlsma, L.; Covaci, A.; Emke, E.; Hernández, F. I.; Reid, M.; Ort, C.; Thomas, K. V.; Van Nuijs, A. L.; De Voogt, P. J. E. s.; technology, Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers. *Environmental Science & Technology*. **2013**, *47*, (3), 1452-1460.
6. Thomas, K. V.; Amador, A.; Baz-Lomba, J. A.; Reid, M. J. Use of mobile device data to better estimate dynamic population size for wastewater-based epidemiology. *Environmental Science & Technology*. **2017**, *51*, (19), 11363-11370.
7. Lai, F. Y.; Ort, C.; Gartner, C.; Carter, S.; Prichard, J.; Kirkbride, P.; Bruno, R.; Hall, W.; Eaglesham, G.; Mueller, J. F. Refining the estimation of illicit drug consumptions from wastewater analysis: Co-analysis of prescription pharmaceuticals and uncertainty assessment. *Water Research*. **2011**, *45*, (15), 4437-4448.
8. Been, F.; Rossi, L.; Ort, C.; Rudaz, S.; Delémont, O.; Esseiva, P. Population Normalization with Ammonium in Wastewater-Based Epidemiology: Application to Illicit Drug Monitoring. *Environmental Science & Technology* **2014**, *48*, (14), 8162-8169.
9. O'Brien, J. W.; Thai, P. K.; Eaglesham, G.; Ort, C.; Scheidegger, A.; Carter, S.; Lai, F. Y.; Mueller, J. F. A model to estimate the population contributing to the wastewater using samples collected on census day. *Environmental Science & Technology*. **2014**, *48*, (1), 517-25.
10. Senta, I.; Gracia-Lor, E.; Borsotti, A.; Zuccato, E.; Castiglioni, S. Wastewater analysis to monitor use of caffeine and nicotine and evaluation of their metabolites as biomarkers for population size assessment. *Water Research*. **2015**, *74*, 23-33.
11. Chen, C.; Kostakis, C.; Gerber, J. P.; Tschärke, B. J.; Irvine, R. J.; White, J. M. Towards finding a population biomarker for wastewater epidemiology studies. *Science of the Total Environment*. **2013**, *487*, 621-628
12. Phung, D.; Mueller, J.; Lai, F. Y.; O'Brien, J.; Dang, N.; Morawska, L.; Thai, P. K. Can wastewater-based epidemiology be used to evaluate the health impact of temperature?—An exploratory study in an Australian population. *Environmental Research*. **2017**, *156*, 113-119.
13. O'Brien, J. W.; Grant, S.; Banks, A. P.; Bruno, R.; Carter, S.; Choi, P. M.; Covaci, A.; Crosbie, N. D.; Gartner, C.; Hall, W. A National Wastewater Monitoring Program for a better understanding of public health: A case study using the Australian Census. *Environment International*. **2019**, *122*, 400-411.
14. The Australian Bureau of Statistics 2900.0 - Census of Population and Housing: Understanding the Census and Census Data, Australia, **2017**; <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2900.0~2016~Main%20Features~Understanding%20Census%20data%20quality~4>, <http://www.webcitation.org/78BXI2Mi5>
15. Harding, S., . Jackson Pulver, L., McDonald, P., Morrison, P., Trewin, D., Voss, A. Report on the Quality of 2016 Census Data, **2017**; <http://www.abs.gov.au/websitedbs/d3310114.nsf/home/Independent+Assurance+Panel/%24File/CIAP+Report+on+the+quality+of+2016+Census+data.pdf>, <http://www.webcitation.org/78BWkiYji>.
16. Doane, D. P.; Seward, L. E. Measuring skewness: a forgotten statistic?. *Journal of Statistics Education*. **2011**, *19*, (2).

17. Tscharke, B.; Mackie, R.; O'Brien, J.; Grant, S.; Mueller, J.; Ghetia, M.; Aghera, H.; Bade, R.; Gerber, C.; White, J. National Wastewater Drug Monitoring Program—Report 4. **2018**; <https://acic.govcms.gov.au/sites/g/files/net3726/f/nwdmp4.pdf?v=1522809564>.
18. Australian Government Official tourist sites, **2019**; <https://www.australia.gov.au/information-and-services/passports-and-travel/official-tourist-sites>, <http://www.webcitation.org/78BaaWRYK>.
19. Spooner, C., Hetherington, K. Social determinants of drug use, **2005**; <https://ndarc.med.unsw.edu.au/sites/default/files/ndarc/resources/TR.228.pdf>.
20. United Nations Department of Economic and Social Affairs, Principles and recommendations for population and housing censuses, **2008**; https://unstats.un.org/unsd/publication/seriesm/seriesm_67rev2e.pdf, <http://www.webcitation.org/78D7qO2cK>.
21. United Nations Statistics Division, 2020 World Population and Housing and Census Programme, **2016**; <https://unstats.un.org/unsd/demographic/sources/census/censusdates.htm>, <http://www.webcitation.org/78D7QH5EI>.