Supplemental Material

Appendix 1: Estimating enteric virus concentrations in raw sludge

One approach for estimating or imputing non-detected values is to use a method called regression on order statistics (ROS). This procedure imputes the non-detected values within the range of 0 to the non-detect level, which for the risk assessment example presented in this report is 1 PFU/4 grams. The details of the method are presented in Shumway et al. (2002) and are shown in Figure A-1. The general idea is that the procedure fits a line to the detected values and then uses this fit to extrapolate down below the detection limit to impute the non-detected values. The sample mean and variance are then calculated from the set of observed and imputed values

This method assumes that the values are normally distributed. Environmental contaminant data are commonly assumed to be log-normally distributed. Based on this assumption, the ROS method is applied to the log-transformed values. The variance of the lognormal distribution of viruses $(\hat{\sigma}_y^2)$ is estimated by calculating the sample variance of the set of observed and estimated values. Transforming the estimates back to the original (non log) scale often introduces transformation bias. To help correct this bias, the Quenouille-Tukey Jackknife method is used to estimate the mean of the lognormal distribution $(\hat{\mu}_y)$ and variance of this mean estimate $(\hat{\sigma}^2_{\hat{\mu}_y})$ (Shumway et al. 2002). Note that the mean estimate is assumed to follow a normal distribution with parameters $\hat{\mu}_y$ and $\hat{\sigma}^2_{\hat{\mu}_y}$. The ROS method is unbiased when the assumption of the lognormal distribution of the untransformed values is correct but is not robust against distributional departures from the lognormal. Therefore it is important to check this assumption.

The results of imputing the non-detect values using ROS for the data shown in Table 1 are provided in Table A-1. The five imputed values range from 0.044 to 0.605. With these imputed values, a complete dataset can be used to estimate pre-treatment concentrations. The pre-treatment concentrations are estimated using a lognormal distribution. The lognormal function requires the identification of two parameters: the sample mean and variance. The estimated mean, $\hat{\mu}_y$, and variance, $\hat{\sigma}_y^2$, are 4.53 PFU/4 grams and 75.5 PFU/4 grams, respectively (see Table A-1). The variance estimate for the lognormal distribution characterizes the variability of the data. There is also uncertainty, which is a function of the sample size (in this case $n = 12$). Assuming that the sampling errors are normally distributed, the estimated variance of the mean concentration, $\hat{\sigma}^2_{\hat{\mu}_y}$, is 6.3.

Figure A-2a shows the lognormal fit for the raw sludge data (the bars represent the actual data). Figure A-2b shows the uncertainty in the mean estimate of the lognormal distribution in Figure A-2a. A left truncated normal distribution was used to describe the mean concentration to avoid obtaining any negative values for the mean estimate. The resulting composite distribution of pretreatment pathogens is highly skewed (Figure A-2c). This skewness is due to extreme values in the distribution. The maximum value is 517, as compared to the mean of 4.6, and there are 72 (0.03%) values greater than 140.

Appendix 2: Simulation Approach

Each simulation is run 25,000 times, each time creating a 1000 kg biosolids pile. For this risk assessment exercise it is assumed that 0.1 g of direct ingestion occurs from each 1000 kg pile. That is, the risk scenario is a biosolids worker ingesting 0.1 g from a 1000 kg biosolids pile. A rotavirus dose-response function is used to estimate risk.

For the solution to converge, 25,000 simulations are required for each scenario. For computational efficiency, therefore, the 1000kg pile is divided into 1000g pieces.

The simulation process for Model1 (repeated 25,000 times for each scenario) is as follows:

1. Randomly sample from a normal distribution with mean $\hat{\mu}_y$ and variance $\hat{\sigma}_{\hat{\mu}_y^2}$ to obtain μ_{cp}, the mean pathogen concentration. The units of μ_{cp} are PFU per 4 grams.
2. Random sample from the lognormal distribution with mean μ_{cp} and variance $\hat{\sigma}_y^2$ to obtain c_p, the actual pathogen concentration. The units of c_p are PFU per 4 grams. Convert the units of c_p into PFU per 1 gram to obtain c_p^*.
3. Randomly sample from an exponential distribution with mean retention time β_{ret} to obtain retention time in digester (t_{ret}). The units of t_{ret} are days.
4. Calculate treatment removal based on line with slope l/β_{ret} where l is a log removal that varies by scenario. Log removal $= t_{ret} l/\beta_{ret}$. Calculate and retain C_p^{**}, the concentration of viruses in the 1000g piece, by attenuating C_p^* using the above calculated log removal.
5. Repeat the above 4 steps 1000 times to accumulate a 1000kg pile.
6. Randomly sample 0.1 gram for the 1000kg biosolids pile and calculate risk using rotavirus dose-response function.

This procedure is repeated 25,000 producing 25,000 estimates that describe the resulting risk distribution.

For the models which include pathways other than direct ingestion, rather than directly sample 100 mg from the biosolids pile as outlined in step 6 above, one begins (at step 6) to apply the model for the exposure pathway in question (groundwater or aerosol, for example). Then the individual or

population-level risk is assessed based on exposure to a given amount of “end product” from the pathway in question.

Appendix 3: Groundwater Model

Up to this point, the risk simulation models were based on direct ingestion from the anaerobic digester. We now consider the scenario where the biosolids are applied to soil, enter the ground-water, and are subsequently ingested from contaminated well water. Viruses from the biosolids are attenuated as they pass through soil, where the degree of attenuation depends on the type of media. In the context of the previous models, this movement through soil can be thought of as additional attenuation of the viruses in the biosolids.

There are three types of groundwater media that are modeled:

1. Non-porous (fractured bedrock or karst). In this case, a fixed 1-3 log-removal is assumed.
2. Unsaturated soil. Attenuation in this case depends on the thickness of the unsaturated layer, as well as other parameters such as its hydraulic conductivity, sorption, and in activation. The Virulo software package (Version 1.0 http://www.epa.gov/ahaazvuc/csmos/models/virulo.html) was used to estimate viral attenuation in various thicknesses of unsaturated soil.
3. Saturated soil. The steady-state solution to a one-site kinetic model was used to estimate viral attenuation in this case. The model depends on attachment and detachment rates, inactivation rates of free and attached viruses, dispersivity, velocity, and distance, the values of which were obtained from a column experiment (Schijven et al. 2002).

Groundwater scenarios may include one or more of these types of media. Fate and transport models for unsaturated and saturated soils were obtained from the literature, as were the distributions of the

corresponding model parameters (see Eisenberg et al for details (2006)). Output distributions of viral concentrations are obtained using Monte Carlo simulation techniques that randomly sample each parameter distribution as well as an input viral concentration distribution estimated using Model 1 or 2.

This exposure model serves to attenuate the virus concentration due to transport of biosolids through the groundwater. These three types of media combined with the many possibilities for barrier depths result in a large number of simulation scenarios. Based on previous simulations of the attenuation in unsaturated soil, a barrier depth of greater than 0.5 meters results in log removals so high as to remove essentially all remaining viruses. Thus, 0.5 meters is the maximum barrier depth used for unsaturated soil. Saturated soil is not as efficient in removing viruses, and thus the maximum barrier depth is 30 meters. The following variations on the groundwater model were intended to capture as much of this variability as possible while not creating an overly complex model.

1. Non-porous (fractured bedrock or karst) media only. For this we assume an attenuation of 1-3 log removal of viruses.
2. Non-porous media followed by 5, 15 or 30 meters saturated soil (3 scenarios).
3. Unsaturated soil followed by saturated soil. In this case there are six possible scenarios: either 0.25 or 0.5 meters of unsaturated soil followed by 5, 15, or 30 meters of saturated soil.

Thus there are 10 groundwater scenarios. Each of these scenarios is run in conjunction with a 1st order anaerobic digester model (Models 1 and 2).

In order to estimate a dose corresponding to the groundwater exposure, several additional assumptions were made. The first assumption is that a rainfall event occurs immediately after the 1000 kg pile of biosolids is applied to land. The second assumption is that a proportion of the viruses from the pile leach from the solids and move down through the soil or karst with the rainwater. The leaching proportion estimate used in the model is 8%. This estimate is based on a study that was aimed at investigating the leaching and transport of viruses. The study found that less than 8% of coliphage initially present in biosolids leached out of the biosolids-soil matrix (Chetochine et al. 2006). The

third assumption is that the well is directly downstream of where the biosolids are applied (Figure A-3). The attenuation in the specific medium (media) for the given scenario is calculated based the downstream distance using either the saturated or unsaturated fate and transport models described above. This attenuation is applied to the remaining 8% of viruses that leached during the rainfall event. The fourth assumption is that the rainwater containing the viruses is diluted with clean water with a ratio of 1:1. The fifth assumption is that the viruses arrive over a period of 3 days, with a third of the viruses arriving each day. The viruses are distributed in the well-pumped water over the three days, based on an average of 150 gallons (567 liters) of water pumped per day. An ingestion rate of 1.2 liters of water per day was used. Based on this dose, the rotavirus dose-response function is used to estimate the single-event risk. The annual risk is also estimated from the model

$$\text{Annual Risk} = 1 - (1 - r)^d$$

where r is the single event risk and d is the number of days exposed. To estimate the annual risk, it is assumed that the biosolids are applied 2 times per year with a 3-day exposure each time for a total of 6 day of exposure per year.

Figure A-3 provides a schematic for the flow of the viruses after a rainfall event for the scenario of unsaturated soil followed by saturated soil. The arrows represent the flow of viruses to the well that is directly downstream of the biosolids pile.

Appendix 4: Aerosol Model

A second possible exposure pathway that is examined is the scenario in which the biosolids are applied to land and the viruses are subsequently transported via aerosolization. An area source model was used to estimate the downwind concentrations of viruses (Eisenberg et al. 2006). Model parameters came from a variety of sources, such as (Brooks et al. 2004; Dowd et al. 2000; Parker et

al. 1977; Pasquill 1962). For details, see Eisenberg et al (2006). The size of the area source plot was assumed to be 100 meters by 100 meters. A 1000 kg pile of biosolids was applied to the given plot. The exposed person breathing the air in each scenario was assumed to be directly downwind of the biosolids application site.

Based on the literature review, the aerosol scenarios include wind speeds of 2, 5, and 10 meters/second at downwind distances of the biosolids application site of 30, 100 and 250 meters. These combinations cover typical meteorological conditions as well as typical distances from biosolids application sites, which range from 30 meters to 300 meters (Personal communication Greg Kester, HAC member). The 9 distinct scenarios were run in conjunction with the first-order digester model.

In order to calculate the dose for the aerosol exposure pathway, the average human breathing rate of 0.83 m^3 per hour, and an exposure duration of 8 hours, were used. Thus the dose was calculated as

$$\text{Concentration in Dose} = c_d \text{ x } 0.83 \text{ x } 8$$

where c_d is the downwind viruses per m^3. The rotavirus dose-response function was used to estimate the individual-level single event risk based on this dose. The annual risk was also estimated from the model. To estimate the annual risk, it was assumed that the biosolids are applied two times per year with a three-day exposure each time for a total of six days of exposure per year.

Appendix 5: Dynamic Model to Estimate Population-Level Risks

For the purposes of this study the community is considered to be open to external sources of pathogens, and infectious members do not shed pathogens in the environment. Thus the shedding parameter was set to 0. For effects of shedding on risk see Eisenberg et al (2004). In addition, the community is not considered homogenous to exposure; i.e., there are occupational exposures to adults and residential exposures to children. The occupational exposure is applied to 1% of the population, and a residential exposure is applied to 50% of the remaining 99% of the population.

Within each group, occupational and residential, the exposure is assumed homogenous.

For the residential exposure, the biosolids are applied two times per year with a three-day exposure each time. For the occupational exposure, workers are exposed five days per week, 52 weeks per year. The model is run for 1,500 days to reach a steady state and subsequently run for another 365 days. The number of disease onsets during the last 365 days is counted resulting in a one-year cumulative incidence. The attributable risk for biosolids (AR) is estimated as the difference in the cumulative risks when biosolids are present and when they are not (Eisenberg et al. 2004).

There are three exposure scenarios for the population-level risk model. The first focuses on occupational exposure only, the second on residential exposure only, and the third on both occupational and residential exposures within a community.

All parameter values, with the exception of shedding, secondary transmission, and the pathogen concentration in the environment due to sources other than biosolids, are set to the values according to the decision tree that classified a high attributable risk, determined in a previous publication (Eisenberg et al. 2004). Other parameters are set to the middle value as determined in (Eisenberg et al. 2004), including those that were either: 1) not factors in the decision tree (see Figure 4 in (Eisenberg et al. 2004)); or 2) not related to shedding, secondary transmission or the pathogen concentration in the environment due to sources other than biosolids. As indicated above, the shedding parameter (φ) is set to 0. The total number of pathogens in the environment (W) is equal to

$$W = W_{ext} + \varphi$$

where

$$W_{ext} = V_b + V_e$$

where V_b is the number pathogens from biosolids and V_e is the number of pathogens from sources other than biosolids. Thus, when shedding is set to 0,

$$W = .V_b + V_e.$$

The estimate of V_e is determined such that when pathogens from biosolids are absent $(V_b = 0)$ and secondary transmission is 0, the incidence is 20 cases per 100,000. The secondary transmission parameter (β_{s2}) is then estimated by the value that results in 40 cases per 100,000 with V_b, V_e, and φ as above. This secondary transmission estimate thus doubles the risk of a composite environmental exposure that excludes biosolids. This particular estimate was not based on data; it was used in this analysis to simply illustrate the potential role of secondary transmission in risk assessment

References

Brooks JP, Gerba CP, Pepper IL. 2004. Biological Aerosol Emission, Fate, and Transport from Municipal and Animal Wastes. Journal of Residuals Science and Technology 1(1):15-28.

Chetochine AS, Brusseau ML, Gerba CP, Pepper IL. 2006. Leaching of phage from Class B biosolids and potential transport through soil. Applied and environmental microbiology 72(1):665-671.

Dowd SE, Gerba CP, Pepper IL, Pillai SD. 2000. Bioaerosol transport modeling and risk assessment in relation to biosolid placement. Journal of Environmental Quality 29(1).

Eisenberg JNS, Moore K, Soller JA, Colford J, Eisenberg D. 2006. Application of a dynamic model to assess microbial health risks associated with beneficial uses of biosolids - Phase II. Water Environment Research Foundation Report 98-REM-1.

Eisenberg JNS, Soller JA, Scott J, Eisenberg D, Colford J. 2004. A Dynamic Model to Assess Microbial Health Risks Associated with Beneficial Uses of Biosolids. Risk Analysis 24(1):221-236.

Parker DT, Spendlove JC, Bondurant JA, Smith JH. 1977. Microbial aerosols from food processing waste spray fields. Journal Water Pollution Control Federation 49(12):2359-2365.

Pasquill F. 1962. Atmospheric diffusion; the dispersion of windborne material from industrial and other sources. London, Princeton, N. J.,: Van Nostrand.

Schijven J, Hassanizadeh S, de Bruin R. 2002. Two-site kinetic modeling of bacteriophages transport through columns of saturated dune sand. Journal Of Contaminant Hydrology 57(3-4):259-279.

Shumway RH, Azari RS, Kayhanian M. 2002. Statistical approaches to estimating mean water quality concentrations with detection limits. Environ Sci Technol 36(15):3345-3353.

Table A-1: Imputed values for nondetects

Month	Raw Data	Imputed Values
1	< 1	**0.044**
2	< 1	**0.124**
3	< 1	**0.234**
4	< 1	**0.388**
5	< 1	**0.605**
6	5	5
7	2	2
8	1	1
9	1	1
10	7	7
11	6	6
12	31	31

Table A-2: List of model assumptions

Modeling enteric virus concentrations in raw sludge

- Occurrence data in raw sludge can be represented using a lognormal distribution; i.e., pathogens in sludge are assumed to be independently and identically distributed. This distribution is widely used in environmental measurements as they take on non-negative values and are frequently skewed to the right.
- Non detectable values are estimated through extrapolation.
- Current analytic methods can only detect a subset of viral pathogens. This together with the fact only a fraction of measureable virus are recovered in any analytic methods and that sampling is limited, indicate that these data should be considered an underestimate of the actual concentrations of viral pathogens present in biosolids.

Modeling the treatment process

- Digesters are well-mixed first order processes and are characterized by a mean log removal rate and a retention time.
- Lime treatment is modeled using a constant attenuation value.

Post treatment concentrations

- Non detectable values from post treatment samples are modeled with a constraint that 99% of all simulated values must be below the detectable limit.
- 1000 Kg. biosolids piles are constructed via simulation.

Biosolids application

- Surface application occurs twice per year. Each application lasts for 3 days.

Exposure

- Direct ingestion exposure occurs either by children that live near application site or workers involved in treatment or application. Children are exposed during application days, while workers are exposed 5 days per week.
- Groundwater exposure occurs by residents that use groundwater as their drinking water source near the application site. Furthermore: 1) A rainfall event occurs immediately after the 1000 kg pile of biosolids is applied to land: 2) A proportion of viruses from the pile (8%) leach from the solids and move down through the soil or karst with the rainwater; 3) The well is directly downstream of where the biosolids is applied; 4) The rainwater containing the viruses is diluted with clean water with a ratio of 1:1; and 5) the viruses arrive over a period of 3 days, with a third of the viruses arriving each day.
- The dose from aerosolized pathogens was assumed to be the product of the concentration at the exposure site (viruses per m3) as predicted by the aerosol model, an average human breathing rate of 0.83 m3 per hour, and an exposure duration of 8 hours. The breathing rate was used to estimate the amount of pathogens entering an individual's mouth; however, to be consistent with the risk estimates from other pathways, it was assumed that those particles were ingested.

Dose response

- A Beta-Poison dose-response relationship was employed.
- The dose-response model was parameterized using data from a rotavirus dosing trial. This is a conservative assumption as rotavirus is thought to be one of the more infectious enteric viruses, whereas the monitoring data is identifying a group of culturable viruses.
- The focal forming units (FFU) used in the dosing trial and the plague form units (PFU) used in the virus occurrence data are equivalent.
- Although the Beta-Poison assumes some level of variability in infectivity, the data that the model is based on come from a relatively healthy homogeneous cohort. Little is known about how the dose response varies for specific subgroups such as the elderly or the immunicompromised.

Population model

- Assumptions for this model are detailed in a previous publication (Eisenberg et al. 2004). Briefly, the model is a deterministic compartmental transmission model that accounts for both environmental transmission, through exposure to pathogens in biosolids and other environmental sources, as well as person-to person transmission, though direct contact with infectious individuals.
- Differences between this and the previously published model are: The community is considered to be open to external sources of pathogens; Infectious members do not shed pathogens in the environment; and Community exposure was split between occupational exposures to adults (1% of population is assumed to be exposed in this manner) and residential exposure to children (50% of population is assumed to be children). For occupational exposure workers are exposed five days per week, 52 weeks per year. For residential exposure children were exposed two times per year with a three-day exposure each event.

Figure A-1: ROS Method for Estimated Censored Values

Estimating the non-detect values. We have $n_0 = 5$ observations, y_i, where $i = 1, 2, , 5$. Each observation is below the detection limit $U = 1$ and $n_1 = 7$ observations are greater than U. Using an appropriate transformation (e.g. $x_i = ln(y_i)$), we assume the observations are independently and normally distributed and have common mean μ_x and variance σ_x^2. The mean and variance will satisfy the equation $x_i = \mu_x + \sigma_x \Phi^{-1}(P_i)$, where P_i is the probability that $X_i < x_i$ ($P_i = \mathbf{P}\{X_i < x_i\}$) and $\Phi^{-1}(P_i)$ is the inverse of the cumulative normal distribution function.

A regression is performed on the inverse order statistics. The procedure is to replace the probabilities by the adjusted ranks so that the regression equation becomes $x_i = \mu_x + \sigma_x \Phi^{-1}(\frac{i-3/8}{n+1/4})$ where $i = n_0+1, n_0+2, \ldots, n_0+n_1$, $n = 12$, and the estimators for μ_x and σ_x are estimated by least squares.

So for the data set used in this report, the 7 observed values are 5, 2, 1, 1, 7, 6, 31 PFU/4 gm (Table 2). The ordered statistics procedure requires that the values are ordered, where the first 5 are non-detects and the following 7 are: 1, 1, 2, 5, 6, 7, 31. These values are next transformed, $x_i = ln(y_i)$. The normal scores for each of the 7 transformed values are estimated using the above function, $\Phi^{-1}(.)$. The resulting 7 values are then used to estimate the two parameters, $\hat{\mu_x}$ and $\hat{\sigma_x^2}$, of the above linear function. The 5 non-detected values $\hat{y_i} = exp(\hat{x_i})$, $i = 1, \ldots, 5$ are then estimated from the same linear model.

Estimating the Variance of the Distribution of Viruses. The variance of the distribution of viruses is estimated by the sample variance $\hat{\sigma_y^2}$ of the set of 12 observed and estimated values.

Estimating the mean. Let $\widetilde{\mu_y}$ represent the sample mean of the set of observed and imputed values transformed back to the original scale. Let $\mu_{[-i]}$ represent the sample mean with the ith observation deleted, where $i = 1, , n$.

Let $\overline{\mu_{[\cdot]}} = n^{-1} \sum_{i=1}^{n} \mu_{[\cdot]}$ be the sample mean of these estimators. The Quenouille-Tukey Jackknife estimator is $\hat{\mu_y} = n\widetilde{\mu_y} - (n-1)\mu_{[\cdot]}$.

Estimating the variance of the mean. The variance of the above Quenouille-Tukey Jackknife estimator is:

$$\hat{\sigma_{\hat{\mu_y}^2}} = \frac{n-1}{n} \sum_{i=1}^{n_1} \left(\mu_{[-i]} \overline{\mu_{[\cdot]}}\right)^2$$

Figure A-2: Pre-treatment concentration of viruses. A) Lognormal distribution of pre-treatment concentrations with imputed values, where solid line is the fitted curve and the bars are the data. B) Normal distribution of the mean value, $\hat{\mu_y}$, of the pre-treatment concentrations. This distribution is left truncated to prevent negative values. C) The composite pre-treatment concentrations based on A and B.

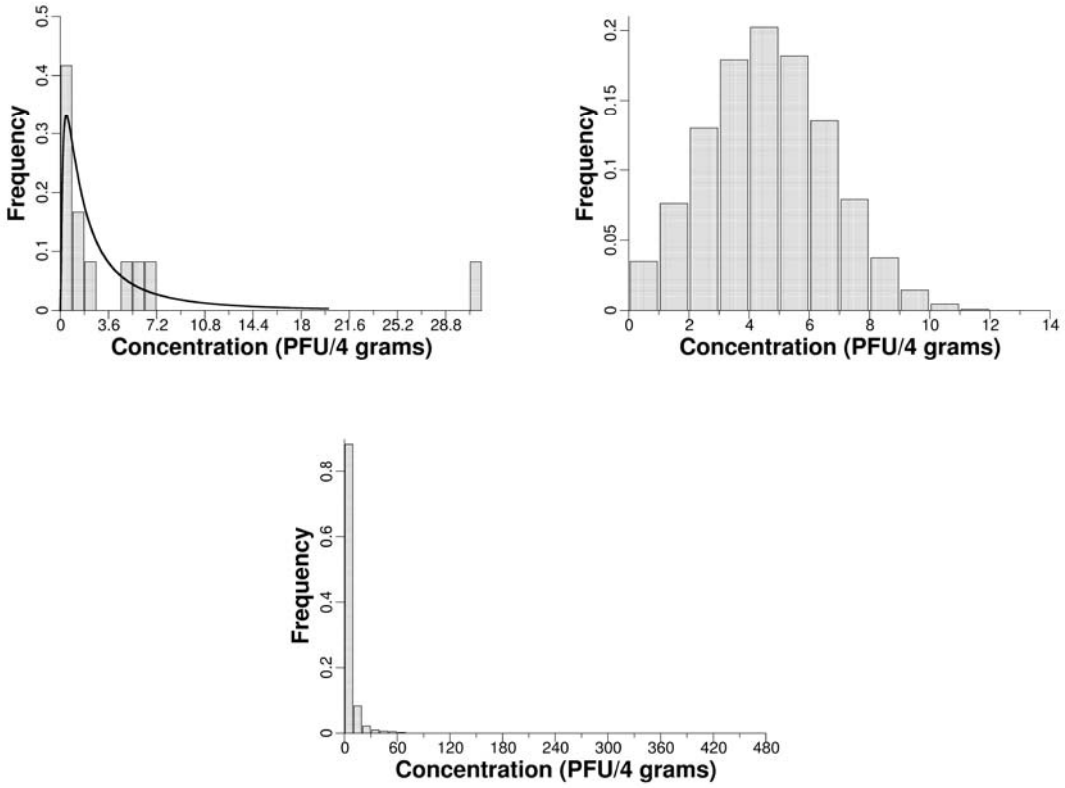

Figure A-3: Groundwater Flow Diagram