Data Appendix
Amanda Kowalski

General Data Description:
One firm in the retail trade industry
2004 and 2003 data
Entire family enrolled from January 1, 2004 to December 31, 2004 plus families with newborns
People who used no care and people with outpatient/inpatient claims
This firm offered four different health plans in 2003 and 2004

Original data format:
Outpatient claims, inpatient claims, and enrollment information available in separate files by year.
Inpatient claims are grouped into admissions in a separate file. I use this file to group inpatient claims
into admissions.
Monthly coverage spells with detailed demographic information are grouped into continuous
enrollment spells in a separate file. I use this file to select on continuous enrollment.
"enrolid" allows linking of claims to enrollment records – never missing on claims files

One-person "families" are included in family-level statistics unless otherwise noted.
I only use the term "individuals" for one-person families.

Selection statistics:
898,512 people in 526,588 families (847,748 people in 491,455 families in 2003) ever observed in the
firm 111 data
898,181 people in 526,290 families (847,550 people in 491,286 families in 2003) after dropping
claims for which I never observe enrollment info for anyone in the family
896,265 people in 525,682 families (845,461 people in 490,543 families in 2003) after dropping
families for which I observe claims for a family member who is never enrolled
881,495 people in 519,703 families (840,548 people in 488,671 families in 2003) after dropping
families for which I observe claims with no corresponding enrollment information
648,154 people in 370,204 families (629,076 people in 355,203 families in 2003) after dropping
families for which a member identified as the employee is enrolled for a spell shorter than Jan
1, 2004 - Dec 31, 2004
573,606 people in 346,858 families (568,559 people in 335,474 families in 2003) after dropping
families for which any member of the family (excluding babies born in 2004 - see below) is
enrolled for a spell shorter than Jan 1, 2004 – Dec 31, 2004
573,530 people in 346,826 families (568,498 people in 335,442 families in 2003) after dropping
families who switched plans in the middle of 2004
573,520 people in 346,821 families (568,495 people in 335,440 families in 2003) after dropping
families with different plans on enrollment vs. claims (Medstat claims there is no authoritative
source)
573,504 people in 346,813 families (568,488 people in 335,437 families in 2003) after dropping
families with different geographic division, state, or hourly vs. salary status on enrollment vs.
claims (There are also discrepancies in "eestatu" which takes on values of "Active Full Time",
"COBRA Continuee", or "Other Unknown" and "Active Full Time" or "COBRA Continuee"
in the enrollment vs. the claims data. I do not drop families with these discrepancies, but they
are dropped in a subsequent step.)

545,897 people in 330,689 families (553,773 people in 326,993 families in 2003) after dropping families who changed geographic division, state, hourly vs. salary status, or "eestatu" discussed above.

543,857 people in 329,669 families (551,415 people in 325,814 families in 2003) after dropping *individuals* who have babies that begin enrollment in the middle of the year (I must drop them because I do not know if family or individual policy parameters apply)

542,289 people in 328,228 families (549,932 people in 324,460 families in 2003) after dropping families in which there is no identifiable employee (I merged the 2003 enrollment data with these claims to see if I could identify any past employees, but I could not)

542,287 people in 328,226 families (549,926 people in 324,456 families in 2003) after dropping families in which the employee is born after 1986 (born after 1985 in 2003) (under age 18 for the entire calendar year).

541,695 people in 327,869 families (548,561 people in 323,700 families in 2003) after dropping families in which the geographic division of the employee cannot be identified.

539,935 people in 326,994 families (533,996 people in 316,391 families in 2003) after dropping families in which the cumulative total payment is more than $1000 different from the sum of the cumulative insurer and beneficiary payments. These families are likely to have claims that have not been reconciled or adjustments that have not been resolved.

537,321 people in 326,051 families (522,898 people in 311,160 families in 2003) that pay more than a dollar over the statutory individual or family maximum deductible in the deductible field. I drop these families because even though I have carefully selected my sample to include only those families without plankey discrepancies, there is a chance that the plan has been misclassified in some of these observations, making the model very wrong. (see Hausman, Abrevaya, Scott Morton 1998).

537,157 people in 325,992 families (522,644 people in 311,050 families in 2003) after dropping families and individuals in which cumulative family or individual pay, coins, or deduct are negative at the end of the year. I drop these individuals because it seems that adjustments have not been resolved for them.

Clean Sample breakdown by family size:
213,735 people (200,814 people in 2003) in families of one.
112,850 people in 56,425 families (107,452 people in 53,726 families in 2003) of two.
78,204 people in 26,068 families (77,532 people in 25,844 families in 2003) of three.
132,368 people in 29,764 families (136,846 people in 30,666 families in 2003) of four or more.

Main estimation sample:
404,789 people in 296,228 (385,798 people in 280,384 families in 2003) families after keeping individuals in families of size three or fewer.
296,228 people after keeping only the employee in each family.
101,343 people after keeping only people who could be matched to median income by zip code.

Data notes:
People born in 2004:
Babies are born at all points in the year, so babies are not always enrolled in health insurance plans for the entire year even if the rest of their families are continuosly enrolled. Accordingly, when I select the sample to include only families in which every member is continuously enrolled, I do not drop families if the only non-continuously enrolled member was born in 2004. I keep these families

because childbirth is a major medical expense, and I would not want to drop most or all families that have babies from my sample. To complicate matters, many babies have claims in the data before they have enrollment records in the data, and I generally drop all families that have claims without corresponding enrollment records. To avoid dropping families with babies, and to make analysis possible, I assume that the coverage information from the baby's first available monthly enrollment spell also applied in all previous months in 2004. If the baby has claims but is never enrolled (which might happen for babies born in December), I drop the family because I never have the baby's enrollment information.

3,367 babies born in 2004 in 3,333 families are in the 2004 clean sample. (4,016 babies born in 2003 in 3,984 families are in the 2003 clean sample.)
3,187 babies in 3,154 families (3,722 babies in 3,690 families in 2003) are enrolled for less than the entire year.
1,831 of these babies in 1,813 families (691 of these babies in 677 families in 2003) have paid claims in the data before they are officially enrolled

Enrollment data discrepancies:
Some people who are continuously enrolled for the entire year according to the summary spell data are missing monthly spells in the detailed spell data, and many people have claims during these months. Medstat acknowledged this discrepancy, and they intend to check for it in future releases of the data. In the meantime, I generated the missing monthly enrollment spells by assuming that the information from the previous month is applicable. There are no gaps longer than one month in the continuously enrolled spells.

6,822 people in 4,574 families (1,311 people in 929 families in 2003) in the clean sample have gaps in their continuous enrollment
383 of these people in 354 families (123 people in 113 families in 2003) have claims during these gaps.

Adjustment records:
Instead of editing claims, changes are made with adjustment records. Medstat resolves many adjustments in the outpatient claims, but they do not resolve adjustments in the inpatient claims. There are financial and non-financial adjustment records in the sample, and they aren't flagged, but I can generally identify them. Both types of adjustments generally have a negative or zero value in the quantity of services "qty" field. Nonfinancial adjustment records have all financial variables set to zero and the corrected values of demographic variables. Financial adjustment records typically have negative values in the financial fields. Whether these negative values result from the "void and replace" method or the "adjustment" method, they should be resolved by adding up claims over time, which I do for my analysis. To resolve non-financial adjustment records, I keep the values of demographic variables from the adjustment records when there is only one other non-adjustment record on the same day. I do not resolve adjustment records if there are many adjustments and claims on the same day. However, I take the median daily value of demographic variables in inpatient and outpatient claims (separately), and I select the sample based on discrepancies between these values and the values in the enrollment file.