STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:605–613
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2039

Ranking states' immunization coverage: an example from the National Immunization Survey[‡]

Lawrence E. Barker[*,†], Philip J. Smith, Robert B. Gerzoff, Elizabeth T. Luman, Mary M. McCauley and Tara W. Strine

Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, U.S.A.

SUMMARY

The National Immunization Survey (NIS) provides state-level estimates of preschool immunization coverage. These coverages are frequently presented in ranked lists, and ranks are frequently over-interpreted. In this paper, we highlight the difficulty in interpreting ranked point estimates. To demonstrate the uncertainty of ranks, parametric bootstrap methods were used to derive 90 per cent confidence intervals for ranks of state vaccination coverage levels among preschool children. We graphically compared states to a reference state. If NIS data are used to rank states, one should consider presenting confidence intervals for rank and the results of comparisons of one state with another graphically. Published in 2005 by John Wiley & Sons, Ltd.

KEY WORDS: immunizations; national immunization survey; parametric bootstrap; ranks

INTRODUCTION

State-level estimates of public health indicators, such as teenage pregnancy, smoking, and rates of violence, are routinely published [1–3]. These estimates aid state health officials in prioritizing needs and developing appropriate public health strategies.

Lists of point estimates are often presented from highest to lowest. Public health professionals, the press, and the public often interpret these as ranks. The simplicity of ranks is appealing. However, most measures in public health result from samples, and the results are presented as point estimates with a specified level of uncertainty. Usually, the 95 per cent confidence intervals (CIs) are presented and, in popular terminology, the CIs' half-widths are called the sample's 'margin of error'. In any ordered list of estimates, translating estimates into ranks compounds uncertainty. However, the uncertainty of ranks is rarely discussed, and has no commonly understood terminology to permit discussion outside research circles.

*Correspondence to: Lawrence E. Barker, Centers for Disease Control and Prevention, 1600 Clifton Road, MS-E62, Atlanta, GA 30333, U.S.A.
†E-mail: lsb8@cdc.gov

Published in 2005 by John Wiley & Sons, Ltd. *Accepted July 2004*

Example from the National Immunization Survey

The National Immunization Survey (NIS) is a large, ongoing, random-digit dialing survey used to provide annual estimates of immunization coverage among 19- to 35-month-old children. In households with an age-eligible child, the respondent is asked for permission to contact the child's immunization provider(s). Later, the surveyed child's immunization providers are asked to submit the child's vaccination record. Provider information is used to determine the number of doses of each vaccine that a child received. The sample is weighted to represent the population of children 19–35 months old during a particular calendar year. Sampling weights account for multiple voice telephone lines in the household, telephone non-response, provider non-response, vital statistics natality data, and non-telephone households. Details of the NIS methods appear elsewhere [4, 5].

The NIS is best suited for estimating nationwide immunization coverage among children aged 19–35 months; at a national level, the standard error for most estimates is less than 0.5 per cent. However, the NIS also provides coverage estimates on a sub-national basis, for example, for states.

When state-level estimates are published, three potentially erroneous interpretations arise. A given state's newest point estimate is often compared with last year's. Such comparisons must be made with awareness of sampling uncertainty [6]. For example, consider coverage with the 4:3:1:3 series (four or more doses of diphtheria and tetanus toxoids and pertussis vaccine, three or more doses of poliovirus vaccine, one or more doses of measles-containing vaccine, and three or more doses of *haemophilus influenzae* type b vaccine) among 19- to 35-month-old children in Arkansas. In 2000, Arkansas' estimated 4:3:1:3 coverage was 72.0 per cent (95 per cent CI, 66.5–77.5 per cent) [7]. In 2001, it was 74.1 per cent (95 per cent CI, 69.4–78.8 per cent) [8]. The difference (2001 coverage rate minus 2000 coverage rate) was 2.1 percentage points (95 per cent CI, −5.1 to +9.3 percentage points). Did the true (as opposed to the estimated) coverage rate increase? Simply put, the survey cannot tell us.

Here is what the CIs tell us: if other samples, of the same size and design, had been taken in Arkansas in 2000 and 2001, about 95 out of 100 resulting CIs for the difference in coverage rates between the 2 years would contain the true, but unobserved, difference in coverage rate. Therefore, we can be 95 per cent confident that the change in coverage rate in Arkansas between 2000 and 2001 was between −5.1 and 9.3 percentage points.

We could calculate narrower CIs, but we would have less confidence that they contained the true value. For a point estimate of coverage, our confidence that it *is* the true value (the implicit assumption made when one does not account for sampling error) is zero. For this example, the apparent increase from 72.0 to 74.1 per cent could be entirely due to sampling uncertainty.

NIS results are often used to compare coverage between states. In the NIS, coverage differences between states are often smaller than the survey's margin of error for those states. For example, in 2001, Maine's estimated 4:3:1:3 coverage was 82.2 per cent (95 per cent CI, 77.7–86.7 per cent), and Maryland's was 77.9 per cent (95 per cent CI, 73.8–82.0 per cent) [8]. The difference (Maine's coverage minus Maryland's) was 4.3 percentage points (95 per cent CI, −1.8 to +10.4 percentage points). Was Maine's true (as opposed to estimated) coverage greater than Maryland's? The NIS cannot tell us. The difference could, with 95 per cent confidence, have been between −1.8 and +10.4 percentage points.

The third potential problem arises when lists of point estimates are translated into ranks. Ranking states from point estimates of coverage introduces even more uncertainty than tracking a state's performance over time, or comparing states' coverage. Nevertheless, ranking is common. One recent news article was headlined: 'Idaho Ranks Last in Child Immunizations' [9]. Small changes in point estimates can cause large changes in rank. For example, in 2001, Georgia was 20th on the list of estimated 4:3:1:3 coverage, at 80.0 per cent (95 per cent CI, 75.7–84.3 per cent) [8]. With a 2 per cent change in point estimate, about one standard error, the state would have moved as low (high) as 27th (11th) on the list.

Here, we quantify uncertainty in states' rankings from the NIS. We also present a graphical method for presenting states' coverage that discourages over-interpretation of results. Those who analyse and report results of other surveys may consider this method of analysing the uncertainty of ranks based on their data, and the possibility of enhancing the presentation of results with easy-to-use graphics.

METHODS

We based this paper on an analysis of state-level 4:3:1:3 coverage in 2001 [8]. We treated the District of Columbia as a state, for 51 state-level estimates. Results would have been similar had we chosen another year, or another measure of immunization coverage (e.g. a single vaccine or another series).

First, we used a parametric bootstrap method to construct the 90 per cent CIs for the rank of each state in 2001 [10]. We chose 90 per cent CIs, instead of the more familiar 95 per cent, to emphasize the great uncertainty in ranks—95 per cent CIs would be even wider. The parametric bootstrap method we used comprises four steps: (1) randomly generating new, hypothetical estimates of coverage for each of the 51 states according to a multivariate normal distribution with mean and standard deviation set equal to the point estimate and the standard error, and with statistical independence among states; (2) ordering those 51 estimates into a ranked list; (3) repeating this process 10 000 times, enough for the results to stabilize; and (4) noting the percentiles of each state's observed ranks. The 5th and 95th percentiles of the ranks define a parametric bootstrap 90 per cent CIs for each state's rank [11].

To better understand how uncertainty of rank varies with the position of a state on a ranked list, we plotted width of CIs (here, we use 'width' to mean 'upper confidence limit minus lower confidence limit', as opposed to 'number of integers contained within the limits') versus rank of point estimate.

A method of graphically comparing a characteristic measured in a reference state to the same characteristic measured in other states has recently been described [12]. It is well known that overlapping CIs do not permit valid comparisons of statistically significant differences [13]. However, intervals which permit valid comparison of multiple states with a reference state via interval overlaps can be calculated [12]. In the method used here, the reference state is represented by a band, defined by $\{(x_{\text{ref}}) \pm k_1(s_{\text{ref}})\}$, and other states are represented by bars defined by $\{(x_i) \pm k_2[(s_{\text{ref}}^2 + s_i^2)^{0.5} - s_{\text{ref}}]\}$, where k_1 and k_2 are appropriate constants, x_{ref} is the point estimate for the reference state, x_i is the point estimate for state i, s_{ref} is the standard error for the reference state, and s_i is the standard error for state i. Details appear elsewhere [12]. Using this method, we can clearly present which states rank ahead of, and behind, a selected reference state, in this case, Georgia.

RESULTS

Table I displays NIS point estimates for 4:3:1:3 coverage rates in 2001, the unweighted number of provider-validated immunizations for each state, the 90 per cent CIs for coverage,

Table I. Ninety per cent confidence limits for rank of state for 4:3:1:3 (four or more diphtheria, tetanus, rubella; three or more polio; one or more measles containing vaccine; three or more *haemophilus influenzae* type b) coverage for 2001, treating the District of Columbia as a state.

Rank of state by point estimate	State (raw sample size)	Estimated coverage (per cent)	90 per cent CI (confidence interval) for coverage (per cent)	90 per cent Bootstrap CI for rank
1	VT (331)	88.0	84.8–91.2	1–5
2	NC (301)	84.7	80.8–88.6	1–17
3	CT (285)	84.1	80.0–88.2	1–19
4	MS (286)	83.9	79.9–87.9	1–20
5	NH (318)	83.9	80.4–87.4	2–18
6	TN (916)	83.9	81.2–86.6	2–15
7	RI (364)	83.7	80.2–87.1	2–18
8	AL (591)	82.7	79.3–86.0	3–22
9	ND (318)	82.5	78.6–86.4	3–24
10	WI (601)	82.5	79.5–85.5	3–21
11	ME (310)	82.2	78.4–86.0	3–25
12	PA (564)	82.0	78.6–85.4	4–24
13	MT (290)	81.7	77.8–85.6	3–27
14	WV (287)	81.0	76.7–85.3	4–30
15	SC (296)	80.8	76.5–85.0	4–31
16	WY (286)	80.6	76.3–84.9	4–32
17	MA (599)	80.6	76.9–84.3	6–30
18	NY (589)	80.5	77.5–83.5	7–28
19	NE (301)	80.4	76.4–84.3	6–32
20	GA (598)	80.0	76.4–83.6	8–32
21	IA (292)	79.4	75.2–83.6	8–36
22	SD (318)	79.1	74.5–83.7	7–39
23	MN (306)	79.0	74.7–83.2	8–38
24	DE (281)	78.6	74.3–82.9	10–39
25	KY (305)	78.5	74.4–82.6	11–39
26	VA (243)	78.0	72.8–83.1	4–30
27	MD (632)	77.9	74.4–81.3	15–38
28	MO (267)	77.8	73.2–82.3	11–42
29	FL (811)	76.9	73.3–80.5	18–43
30	OK (315)	76.2	71.7–80.7	17–47
31	NJ (636)	76.2	71.7–80.7	17–47
32	KS (269)	75.7	69.9–81.5	14–49
33	IL (602)	75.6	72.1–79.0	23–46
34	WA (592)	75.5	71.9–79.1	22–46
35	CO (368)	75.4	71.5–79.2	22–47
36	CA (1261)	74.9	71.9–77.9	26–46
37	OH (920)	74.7	71.3–87.0	26–48
38	DC (311)	74.2	69.5–78.9	23–50
39	UT (302)	74.1	69.4–78.8	24–50

Table I. *Continued.*

Rank of state by point estimate	State (raw sample size)	Estimated coverage (per cent)	90 per cent CI (confidence interval) for coverage (per cent)	90 per cent Bootstrap CI for rank
40	AK (289)	74.1	69.5–78.7	24–50
41	ID (327)	74.1	69.7–78.5	24–49
42	AR (391)	74.1	70.1–78.0	26–49
43	MI (601)	73.9	69.7–78.1	26–50
44	TX (1594)	73.7	70.5–76.9	30–49
45	IN (559)	73.6	69.6–77.5	28–49
46	OR (286)	73.0	68.1–77.9	26–51
47	AZ (590)	72.9	69.4–76.3	32–50
48	HI (278)	72.8	67.2–78.3	25–51
49	NV (297)	72.2	67.4–77.0	30–51
50	NM (338)	71.0	66.7–75.3	35–51
51	LA (572)	68.9	64.2–73.6	42–51

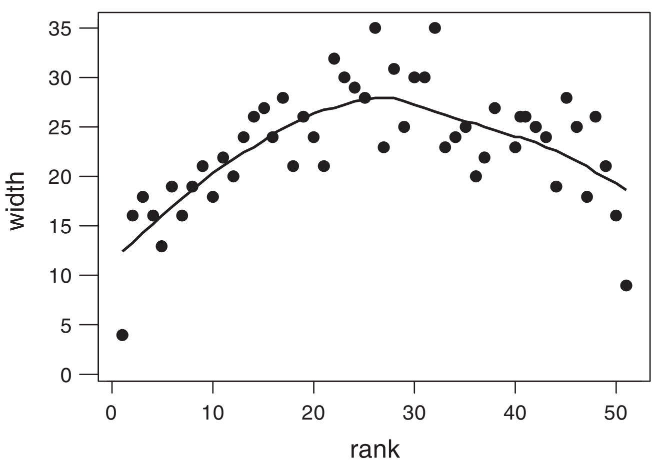

Figure 1. Scatter plot of width of 90 per cent confidence limits for rank versus rank of point estimate of 4:3:1:3 (at least four doses of diphtheria/tetanus/pertussis, at least three doses of polio, at least one dose of measles containing vaccine, at least three doses of *haemophilus influenzae* type b vaccine) coverage. A lowess smooth is included to ease visual interpretation.

and the apparent ranks from simply ordering by point estimate. The table also displays the parametric bootstrap 90 per cent CIs for each state's rank. The CIs demonstrate that state ranks are subject to tremendous uncertainty. In the most extreme examples, Kansas could (with 90 per cent confidence) rank from 14th to 49th. In the least extreme, Vermont could rank 1st–5th. Over the 51 states, the median width for CIs was 24.

For all but the few highest and lowest ranked states, CIs for rank are wide. Figure 1 presents a scatter plot and lowess smooth (a graphical method of presenting a smooth picture of a scatter plot) of width of CI versus rank of point estimate [14]. Figure 1 illustrates that

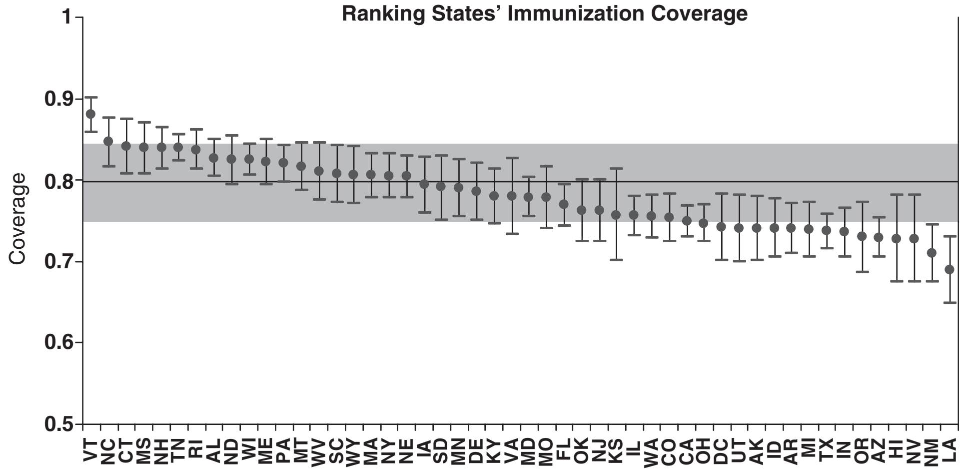

Figure 2. Graphical comparison of 4:3:1:3 (at least four doses of diphtheria/tetanus/pertussis, at least three doses of polio, at least one dose of measles containing vaccine, at least three doses of *haemophilus influenzae* type b vaccine) coverage in the state of Georgia with that of the other states. States with bars that overlap the coloured vertical bar have 4:3:1 coverage which is not statistically significantly different from that of Georgia.

estimated ranks in the middle of the list tend to have wider CIs for rank. This occurs because ranks cannot be better than 1 or worse than 51; states with point estimates of ranks near the top and the bottom cannot possibly have true ranks that are much more extreme in the same direction, but the true ranks might well be more toward the middle.

Depicting point estimates and CIs graphically communicates the difficulty of interpreting a state's point estimate of coverage by simply translating it to a rank (Figure 2). For Georgia, the reference state in the figure, the point estimate of coverage is depicted by the horizontal line, and the shading represents the confidence intervals. Other states appear along the horizontal axis, ordered by their point estimates of coverage, and appropriate intervals about those estimates are indicated. In Figure 2, intervals are defined so that a given state's coverage differs from Georgia's, at the 0.10 level, if and only if the intervals overlap.

The figure illustrates the limits of our ability to interpret these point estimates as ranks. We can say with 90 per cent confidence that, in 2001, one state (Vermont) had higher coverage than Georgia, that two states (New Mexico and Louisiana) had lower coverage than Georgia. Based on NIS data, we cannot distinguish coverage in Georgia from that in the remaining states. This is a different method of analysis than the parametric bootstrap discussed earlier; it is not surprising that it results in different estimates of uncertainty in ranks.

DISCUSSION

Uncertainty in ranks of states based on 4:3:1:3 coverage measured by the 2001 NIS can be quite large, particularly for states other than those with the highest- and lowest-ranking point estimates. This would be true for any survey in which the state-level margins of error were greater than the difference in consecutively ranked states. Thus, ranks should be interpreted with extreme caution, if they are used at all.

In surveys such as the NIS, an analysis of the CIs for rank is informative. For example, Table I shows that four states (Connecticut, North Carolina, Mississippi, and Vermont) could, with 90 per cent confidence, have had the largest 4:3:1:3 coverage in 2001. To acknowledge data limitations, we would simultaneously recognize the states of list when recognizing the 'top' state; this could be further refined by indicating the how many realizations made each state was 'top'. Similarly, to recognize the 'top five' states, as public health officials sometimes do, we should acknowledge all 17 states whose rank might plausibly have been five or smaller (Alabama, Connecticut, Maine, Mississippi, Montana, New Hampshire, North Carolina, North Dakota, Pennsylvania, Rhode Island, South Carolina, Tennessee, Vermont, Virginia, West Virginia, Wisconsin, and Wyoming). Virginia, whose point estimate is 26th, is on this list because it could, due to sampling uncertainty, be in the 'top five'.

If state level estimates of the characteristic of interest can be reasonably treated as normally distributed, parametric bootstrap analysis of ranks requires only point estimates and standard deviations. The graphical method presented here provides a readily computed and easily understood summary of the uncertainty of the rank of a given state. In addition, this method presents a simple visual test for the size of the difference between the reference state and other states. When results of state-based surveys are communicated to state programs, individual graphs with each state as reference could easily be provided.

More precise estimates of coverage would result in more precise ranking based on point estimates. However, coverage is relatively good in most states, and more precise point estimates and ranks would not likely result in improved public health interventions. Thus, reducing sampling uncertainty for a large sample like the NIS through increased sample size is unjustifiably expensive. However, states' standing can be compared to others' using the visual depiction described here. Producing such graphs for each state is not prohibitively time-consuming, nor does it overburden information dissemination channels.

For most data needed in public health, such as immunization coverage, a census of all persons will likely never be available. Thus, the limitations of surveys will continue to apply. Communicating results of surveys in the most meaningful and applicable way, along with their benefits and limitations, will remain crucial.

The number of observations in each state is large enough so that normality is a reasonable approximation to the distribution of the state estimates. Other approaches, such as the use of other distributions or a non-parametric bootstrap, are possible.

Beta distributions might have been used for the estimated coverages. The practical consequences would have been small. In our simulations, Vermont's coverage was fitted as a normal distribution with mean 0.88 and standard deviation 0.02. Using the method of moments (and carrying the mean and standard deviation to one more significant digit than reported here), the beta (245.84, 33.54) distribution is the best fitting beta distribution for Vermont's coverage. The sample size for Vermont was not exceptionally large. Vermont had the point estimate of coverage furthest from 0.5. Therefore, Vermont's estimated coverage should be least normally distributed of any state's. Figure 3 displays the overlaid fitted normal and beta cumulative distribution functions; the two are almost identical.

In theory, a parametric bootstrap assuming a normal distribution could result in a simulated coverage of less than 0 or more than 100 per cent. This did not happen. To see why, we consider Vermont, the state with the point estimate of coverage (88.0 per cent) closest to

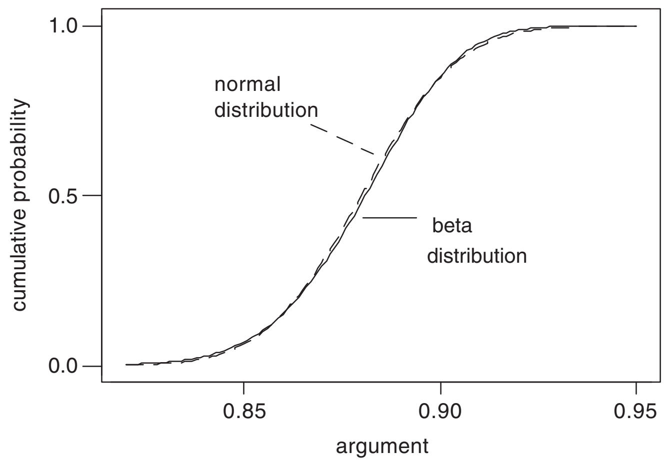

Figure 3. Graphical comparison of beta (245.84, 33.52) and a normal (0.88, 0.02) cumulative distribution functions, the beta and normal distributions which best fit coverage for Vermont. This illustrates how the use of beta, rather than normal, distributions would have changed little in the simulations.

100 per cent. (Louisiana, the 'rank 51' state, had coverage more than 28 standard deviations from zero.) Vermont's coverage was more than six standard deviations away from 100 per cent. The probability that a single realization for Vermont would exceed 100 per cent is approximately 3.6×10^{-10}. Corresponding probabilities for other states would be even smaller (example: the coverage of Virginia, the 'rank 26' state, was 7.1 standard deviations from 100 per cent). Thus, if we assume that, for all 51 states, the probability of a realization exceeding 100 per cent is 3.6×10^{-10}, we overstate the probability of a simulated coverage of more than 100 per cent. More than 500 000 realizations (we used 10 000) would be required to have a 0.01 probability of even one simulated coverage in even one state exceeding 100 per cent.

The non-parametric bootstrap, in which observations are randomly resampled from the data, could have been used. Non-parametric bootstrap analyses of data from the NIS has been explored [15]; the NIS' highly variable sampling weights (in 2001, child-level weights ranged from 3.9 to 3 731.1) and highly complex survey design makes solutions of problems via the non-parametric bootstrap require tens of hours of desktop computing time. In contrast, the non-parametric bootstrap requires almost no time, and produces almost identical answers.

Table I presents CIs for each state, considered separately from any other state. One might wish to consider several states simultaneously. The methods used here could be modified in a fairly obvious manner if one wished to simultaneously consider several states' ranks; this accounting for multiple comparisons would result in even greater sampling uncertainty. We have not pursued this here.

Finally, alternate methods of ranking might be considered. For example, one might rank each state on each antigen and then base a 'rank' on the arithmetic or geometric mean of these ranks. Additional analyses, not reported here, indicate that the sampling uncertainty from such methods is no less than, and in some cases greater than, the uncertainty associated with simple ranking based on point estimates.

ACKNOWLEDGEMENTS

The authors wish to thank Susan Y. Chu, Ph.D., of the Centers for Disease Control and Prevention for her help in finding a direction for this paper, and Patricia Smith, also of the Centers for Disease Control and Prevention, for assistance with graphics.

REFERENCES

1. The Alan Guttmacher Institute. *Teenage Pregnancy*: *Overall Trends and State by State Information*, 1999. www/agi-usa.org/pubs/teen preg stats
2. Centers for Disease Control and Prevention. *Current Cigarette Smoking Among Adults Aged 18 and Older*, 1997. www.cdc.gov/tobacco/statehi/pdf/adultuse
3. National Criminal Justice Commission. *Report Card on Safety*, 2002. www.igc.org/ncia/REPC.HTML
4. Zell E, Ezzati-Rice T, Battaglia M, Wright R. National Immunization Survey: the methodology of a vaccination surveillance system. *Public Health Reports* 2000; **15**(1):65–77.
5. Smith PJ, Battaglia MP, Huggins VJ, Hoaglin DC, Roden AS, Khare M, Ezzati-Rice TM, Wright RW. Overview of the sampling design and statistical methods used in the National Immunization Survey. *American Journal of Preventive Medicine* 2001; **20**(4S):61–68.
6. Simpson DM, Rodewald LE, Barker LE. What's in a number? The use and abuse of survey data. *American Journal of Preventive Medicine* 2001; **20**(4S):86–87.
7. Centers for Disease Control and Prevention. National, state, and urban area vaccination coverage levels among children aged 19–35 months—United States, 2000. *Morbidity and Mortality Weekly Report* 2001; **50**(30): 637–641.
8. Centers for Disease Control and Prevention. National, state, and urban area vaccination coverage levels among children aged 19–35 months—United States, 2001. *Morbidity and Mortality Weekly Report* 2002; **51**(30): 664–666.
9. Bow B. Idaho ranks last in child immunizations. *Spokesman-Review*. Associated Press: Spokane, WA, 18 April, 2001.
10. Gerzoff RB, Williamson GD. Who's number one? *Public Health Reports* 2001; **116**(2):1–12.
11. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall: New York, 1993.
12. Almond RG, Lewis C, Tukey JW, Yan D. Displays for comparing a given state to many others. *American Statistician* 2000; **54**:89–93.
13. Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. *American Statistician* 2001; **55**:182–186.
14. Cleveland WS. *Visualizing Data.* Hobart Press: Summit, NJ, 1993.
15. Smith P, Hoaglin D, Battaglia M, Barker L. Implementation and applications of bootstrap methods for the National Immunization Survey. *Statistics in Medicine* 2003; **22**:2487–2502.