

Selecting the Minimum Denominator in Manual and Enumerated CTA

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

Two methods for specifying the minimum endpoint sample size for optimal (maximum-accuracy) classification tree analysis (CTA) models are noted.

Hierarchically (also known as “manual” or “algorithmic”) optimal (maximum-accuracy) classification tree analysis (CTA), and its evolutionary development—enumerated (also known as “automated”) CTA—have been used to study a wide domain of phenomena and have identified accurate multivariable models.¹⁻⁷⁸

All CTA models terminate in endpoints representing sample strata.⁷⁹ UniODA, Mega-ODA, and CTA software—the only software systems capable of identifying maximum-accuracy CTA models—allow the user to specify the minimum sample size for any strata identified by the model.⁸⁰⁻⁸⁵ This note discusses two methods for determining the appropriate minimum endpoint sample size.

The first method is to conduct a statistical power analysis. This may be accomplished using the results of a simulation study⁸⁶ or, for a “worst-case” (conservative) analysis, by conducting a power analysis using Fisher’s exact test—for which Type I error is isomorphic with a parallel UniODA analysis involving purely binary data.⁸⁰ For large samples this approach will yield models that are not parsimonious.⁷⁹ The second method is a heuristic approach of setting the minimum strata

size to 10% of the overall sample N, in order to identify a model that is capable of being replicated using a hold-out sample reflecting a 90% reduction in overall N.

The final revolution in CTA modeling, novometric theory, employs methods that explicitly identify the globally-optimal model for any sample, but commercial software for conducting novometry is not yet available.^{79,87,88}

References

¹Albuquerque, K., Giangreco, D., Morrison, C., Siddiqui, M., Sinacore, J., Potkul, R., & Roeske, J. (2011). Radiation-related predictors of hematologic toxicity after concurrent chemoradiation for cervical cancer and implications for bone marrow-sparing pelvic IMRT. *International Journal of Radiation Oncology * Biology * Physics*, 79, 1043-1047.

²Alshehlee, A., Ranawat, N., Syed T.U., Conway D., Ahmad, S.A., & Zaiday, O.O. (2010). National Institutes of Health Stroke Scale assists in predicting the need for percutaneous endoscopic gastrostomy tube placement

in acute ischemic stroke. *Journal of Stroke and Cerebrovascular Diseases*, 19, 347-352.

³Arozullah, A.M., Gordon, H.S., Yarnold, P., Soltysik, R., Ferreira, M.R., Wolf, M.S., Molokie, R., Bhoopalani, N., & Bennett, C.L. (2008). Predictors of prostate cancer stage at presentation. *Journal of General Internal Medicine*, 23, 376.

⁴Arozullah, A.M., Lee, S.D., Khan, T., Kurup, S., Ryan, J., Bonner, M., Soltysik, R.C., & Yarnold, P.R. (2006). The roles of low literacy and social support in predicting the preventability of hospital admission. *Journal of General Internal Medicine*, 21, 140-145.

⁵Arozullah, A.M., Parada, J., Bennett, C.L., Deloria-Knoll, M., Chmiel, J.S., Phan, L., & Yarnold, P.R. (2003). A rapid staging system for predicting mortality from HIV-associated community-acquired pneumonia. *Chest*, 123, 1151-1160.

⁶Arozullah, A.M., Yarnold, P.R., Weinstein, R.A., Nwadiaro, N., McIlraith, T.B., Chmiel, J.S., Sipler, A.M., Chan, C., Goetz, M.B., Schwartz, D., & Bennett, C.L. (2000). A new preadmission staging system for predicting inpatient mortality from HIV-associated *Pneumocystis carinii* pneumonia in the early-HAART era. *American Journal of Respiratory and Critical Care Medicine*, 161, 1081-1086.

⁷Belknap, S.M., Moore, H., Lanzotti, S.A., Yarnold, P.R., Getz, M., Deitrick, D.L., Peterson, A., Akesson, J., Maurer, T., Soltysik, R.C., & Storm, J. (2008). Application of software design principles and debugging methods to an analgesia prescription reduces risk of severe injury from medical use of opioids. *Clinical Pharmacology and Therapeutics*, 84, 385-392.

⁸Belmares, J., Gerding, D.N., Parada, J.P., Miskevics, S., Weaver, F., & Johnson, S. (2007). Outcome of metronidazole therapy for *Clostridium difficile* disease and correlation with a scoring system. *Journal of Infection*, 55, 495-501.

⁹Brocklehurst, P.R., Baker, S.R., & Speight, P.M. (2010). Factors which determine the referral of potentially malignant disorders by primary care dentists. *Journal of Dentistry*, 38, 569-578.

¹⁰Brocklehurst, P.R., Baker, S.R., & Speight, P.M. (2010). Primary care clinicians and the detection and referral of potentially malignant disorders of the mouth: A summary of the current evidence. *Primary Dental Care*, 17, 65-71.

¹¹Bryant, F.B. (2010). The Loyola experience (1993-2009): Optimal data analysis in the Department of Psychology. *Optimal Data Analysis*, 1, 4-9.

¹²Bryant, F.B., & Yarnold, P.R. (2014). Finding joy in the past, present, and future: The relationship between Type A behavior and savoring beliefs among college undergraduates. *Optimal Data Analysis*, 3, 36-41.

¹³Bryant, F.B., & Yarnold, P.R. (2014). Type A behavior, pessimism and optimism among college undergraduates. *Optimal Data Analysis*, 3, 32-35.

¹⁴Coakley, R.M., Holmbeck, G.N., & Bryant, F.B. (2006). Constructing a prospective model of psychosocial adaptation in young adolescents with spina bifida: An application of optimal data analysis. *Journal of Pediatric Psychology*, 31, 1084-1099.

¹⁵Coakley, R.M., Holmbeck, G.N., Bryant, F.B., & Yarnold, P.R. (2010). Manual vs. automated

CTA: Predicting adolescent psychosocial adaptation. *Optimal Data Analysis, 1*, 55-58.

¹⁶Cohen, R., Wiley, S., Oswald, D.P., Eakin, K.B., Best, A.I.M. (1999). Applying utilization management principles to a comprehensive service system for children with emotional and behavioral disorders and their families: A feasibility study. *Journal of Child and Family Studies, 8*, 463-476.

¹⁷Collinge, W., & Yarnold, P.R. (2001). Transformational breath work in medical illness: Clinical applications and evidence of immunoenhancement. *Subtle Energies & Energy Medicine, 12*, 139-156.

¹⁸Collinge, W., Kahn, J., Walton, T., Kozak, L., Bauer-Wu, S., Fletcher, K., Yarnold, P.R., & Soltysik, R.C. (2013). Touch, caring, and cancer: Randomized controlled trial of a multimedia caregiver education program. *Supportive Care in Cancer, 21*, 1405-1414.

¹⁹Collinge, W., Yarnold, P., & Soltysik, R. (2013). Fibromyalgia symptom reduction by online behavioral self-monitoring, longitudinal single subject analysis and automated delivery of individualized guidance. *North American Journal of Medical Sciences, 5*, 546-553.

²⁰Collinge, W., Yarnold, P.R., & Raskin, E. (1998). Use of mind/body self-healing practice predicts positive health transition in chronic fatigue syndrome: A controlled study. *Subtle Energies & Energy Medicine, 9*, 171-190.

²¹Collinge, W.C., Soltysik, R.C., & Yarnold, P.R. (2010). An internet-based intervention for fibromyalgia self-management: Initial design and alpha test. *Optimal Data Analysis, 1*, 163-175.

²²Cromley, T., & Lavigne, J.V. (2008). Predictors and consequences of early gains in

child psychotherapy. *Psychotherapy: Theory, Research, Practice, Training, 45*, 42-60.

²³Donenberg, G.R., Bryant, F.B., Emerson, E., Wilson, H.W., & Pasch, K.E. (2003). Tracing the roots of early sexual debut among adolescents in psychiatric care. *Journal of the American Academy of Psychiatry, 42*, 594-608.

²⁴Dunleavy, A.M., & Leon, S.C. (2011). Predictors for resolution of antisocial behavior among foster care youth receiving community-based services. *Children and Youth Services Review, 33*, 2347-2354.

²⁵Feinglass, J., Yarnold, P.R., Martin, G.J., & McCarthy, W.J. (1998). A classification tree analysis of selection for discretionary treatment. *Medical Care, 36*, 740-747.

²⁶Frazier, T.W., Youngstrom, E.A., Fristad, M.A., Demeter, C., Birmaher, B., Kowatch, R.A., Arnold, L.E., Axelson, D., Gill, M.K., Horwitz, S.M., & Findling, R.L. (2014). Improving clinical prediction of bipolar spectrum disorders in youth. *Journal of Clinical Medicine, 3*, 218-232.

²⁷Green, D., Hartwig, D., Chen, D., Soltysik, R.C., & Yarnold, P.R. (2003). Spinal cord injury risk assessment for thromboembolism (SPIRATE Study). *American Journal of Physical Medicine and Rehabilitation, 12*, 950-956.

²⁸Greenleaf, R.G., Flexon, J.L., Lurigio, A.J., & Snowden, J.A. (2010). Predicting injuries of women in episodes of intimate partner violence: Individual and composite risk factors. *Victims & offenders: An International Journal of Evidence-based Research, Policy, and Practice, 5*, 101-119.

²⁹Grobman, W.A., Terkildsen, M.F., Soltysik, R.C., & Yarnold, P.R. (2008). Predicting

outcome after emergent cerclage using classification tree analysis. *American Journal of Perinatology*, 25, 443-448.

³⁰Han, S.D., Suzuki, H., Drake, A.I., Jak, A.J., Houston, W.S., & Bondi, M.W. (2009). Clinical, cognitive, and genetic predictors of change in job status following traumatic brain injury in a military population. *Journal of Head Trauma Rehabilitation*, 24, 57-64.

³¹Han, S.D., Suzuki, H., Jak, A.J., Chang, Y.L., Salmon, D.P., & Bondi, M.W. (2010). Hierarchical cognitive and psychosocial predictors of amnesic mild cognitive impairment. *Journal of the International Neuropsychological Society*, 16, 721-729.

³²Hill RM, Pettit JW, Lewinson PM, Seeley JR, Klein DN (2014). Escalation to major depressive disorder among adolescents with subthreshold depressive symptoms: Evidence of distinct subgroups at risk. *Journal of Affective Disorders*, 158, 133-138.

³³Jones, A., & Ingram, M.V. (2011). A comparison of selected MMPI-2 and MMPI-2-RF validity scales in assessing effort on cognitive tests in a military sample. *The Clinical Neurologist*, 7, 1207-1227.

³⁴Kanter, A.S., Spencer, D.C., Steinberg, M.H., Soltysik, R.C., Yarnold, P.R., & Graham, N.M. (1999). Supplemental vitamin B and progression to AIDS and death in black South African patients infected with HIV. *Journal of Acquired Immune Deficiency Syndromes*, 21, 252-253.

³⁵Kim, B., Lyons, T.M., Parada, J.P., Uphold, C.R., Yarnold, P.R., Hounshell, J.B., Sipler, A.M., Goetz, M.B., DeHovitz, J.A., Weinstein, R.A., Campo, R.E., & Bennett, C.L. (2001). HIV-related *Pneumocystis carinii* pneumonia in older patients hospitalized in the early HAART

era. *Journal of General Internal Medicine*, 16, 583-589.

³⁶Kyriacou, D.N., Yarnold, P.R., Stein, A.C., Schmitt, B.P., Soltysik, R.C., Nelson, R.R., Frerichs, R.R., Noskin, G.A., Belknap, S.B., & Bennett, C.L. (2007). Discriminating inhalational anthrax from community-acquired pneumonia using chest radiograph findings and a clinical algorithm. *Chest*, 131, 489-495.

³⁷Kyriacou, D.M., Yarnold, P.R., Soltysik, R.C., Wunderink, R.G., Schmitt, B.P., Parada, J.P., & Adams, J.G. (2008). Derivation of a triage algorithm for chest radiography of community-acquired pneumonia in the emergency department. *Academic Emergency Medicine*, 15, 40-44.

³⁸Lavigne, J.V., LeBailly, S.A., Gouze, K.R., Binns, H.J., Keller, J., & Pate, L. (2010). Predictors and correlates of completing behavioral parent training for the treatment of oppositional defiant disorder in pediatric primary care. *Behavior Therapy*, 41, 198-211.

³⁹Layden, B.T., Minadeo, N., Suhy, J., Abukhdeir, A.M., Metreger, T., Foley, K., Borge, G., Crayton, J.W., Bryant, F.B., & Mota de Freitas, D. (2004). Biochemical and psychiatric predictors of Li⁺ response and toxicity in Li⁺-treated bipolar patients. *Bipolar Disorders*, 6, 53-61.

⁴⁰Lyons, A.M., Leon, S.C., Zaddach, C., Luboyeski, E.J., & Richards, M. (2009). Predictors of Clinically Significant Sexual Concerns in a Child Welfare Population. *Journal of Child and Adolescent Trauma*, 2, 28-45.

⁴¹Lyons, J.S. (1997). The evolving role of outcomes in managed health care. *Journal of Child and Family Studies*, 6, 1-8.

- ⁴²Martin J, Benjamin E, Craver C, Kroch E, Nelson E, Bankowitz R (2014). Measuring adverse events in hospitalized patients: An administrative method for measuring harm. *Journal of Patient Safety*, April 8, doi: 10.1097/PTS.0000000000000070
- ⁴³Millis, S.R., Ross, S.R., & Ricker, J.H. (1998). Detection of incomplete effort on the Wechsler Adult Intelligence Scale-Revised: A cross-validation. *Journal of Clinical and Experimental Neuropsychology*, 20, 167-173.
- ⁴⁴Mueser, K.T., Yarnold, P.R., Rosenberg, S.D., Drake, R.E., Swett, C., Miles, K.M., & Hill, D. (2000). Substance use disorder in hospitalized severely mentally ill psychiatric patients: Prevalence, correlates, and sub-groups. *Schizophrenia Bulletin*, 26, 179-193.
- ⁴⁵Nebeker, J.R., Yarnold, P.R., Soltysik, R.C., Sauer, B.C., Sims, S.A., Samore, M.H., Rupper, R.W., Swanson, K.M., Savitz, L.A., Shinogle, J., & Xu, W. (2007). Developing indicators of inpatient adverse drug events through non-linear analysis using administrative data. *Medical Care*, 45, S81-S88.
- ⁴⁶Ostrander, R., Weinfurt, K.P., Yarnold, P.R., & August, G. (1998). Diagnosing attention deficit disorders via the BASC and the CBCL: Test and construct validity analyses using optimal discriminant classification trees. *Journal of Consulting and Clinical Psychology*, 66, 660-672.
- ⁴⁷Pape, T.L.B., Guernon, A., Lundgren, S., Patil, V., Herrold, A.A., Smith, B., Blahnik, M., Picon, L.M., Harton, B., Peterson, M., Mallinson, T., & Hoffman, M. (2013). Predicting levels of independence with expressing needs and ideas 1 year after severe brain injury. *Rehabilitation Psychology*, 58, 253-262.
- ⁴⁸Rhode, P., Stice, E., & Gau, J.M. (2012). Effects of three depression prevention interventions on risk for depressive disorder onset in the context of depression risk factors. *Prevention Science*, 13, 584-593.
- ⁴⁹Rupert, P.A., Miller, A.O., Tuminello-Hartman, E.R., & Bryant, F.B. (2012). Predictors of career satisfaction among practicing psychologists. *Professional Psychology: Research and Practice*, 43, 495-502.
- ⁵⁰Sieracki, J.H., Fuller, A.K., Leon, S.C., Jhe Bai, G., & Bryant, F.B. (2015). The role of race, socioeconomic status, and System of Care services in placement decision-making. *Children and Youth Services Review*, doi: 10.1016/j.childyouth.2014.12.013
- ⁵¹Smart, C.M., Nelson, N.W., Sweet, J.J., Bryant, F.B., Berry, D.T.R., Granacher, R.P., & Heilbronner, R.L. (2008). Use of MMPI-2 to predict cognitive effort: A hierarchically optimal classification tree analysis. *Journal of the International Neuropsychological Society*, 14, 842-852.
- ⁵²Smith, J.H., Bryant, F.B., Njus, D., & Posavac, E.J. (2010). Here today, gone tomorrow: Understanding freshman attrition using Person-Environment Fit Theory. *Optimal Data Analysis*, 1, 101-124.
- ⁵³Snowden, J., Leon, S., & Sieracki, J. (2008). Predictors of children in foster care being adopted: A classification tree analysis. *Children and Youth Services Review*, 30, 1318-1327.
- ⁵⁴Snowden, J.A., Leon, S.C., Bryant, F.B., & Lyons, J.S. (2007). Evaluating psychiatric hospital admission decisions for children in foster care: An optimal classification tree analysis. *Journal of Child and Adolescent Psychology*, 36, 8-18.
- ⁵⁵Soltysik, R.C., & Yarnold, P.R. (2010). The use of unconfounded climatic data improves

atmospheric prediction. *Optimal Data Analysis, 1*, 67-100.

⁵⁶Soltysik, R.C., & Yarnold, P.R. (2014). Hierarchically optimal classification tree analysis of adverse drug reactions secondary to warfarin therapy. *Optimal Data Analysis, 3*, 23-24.

⁵⁷Stalans, L.J., & Finn, M.A. (1995). How novice and experienced officers interpret wife assaults: Normative and efficiency frames. *Law & Society Review, 29*, 301-335.

⁵⁸Stalans, L.J., & Seng, M. (2006). Identifying subgroups at high risk of dropping out of domestic batterer treatment: The buffering effects of a high school education. *International Journal of Offender Therapy and Comparative Criminology, 10*, 1-19.

⁵⁹Stalans, L.J., Hacker, R., & Talbot, M.E. (2010). Comparing nonviolent, other-violent, and domestic batterer sex offenders: Predictive accuracy of risk assessments on sexual recidivism. *Criminal Justice and Behavior, 37*, 613-628.

⁶⁰Stoner, A.M., Leon, S.C., & Fuller, A.K. (2013). Predictors of reduction in symptoms of depression for children and adolescents in foster care. *Journal of Child and Family Studies, 22*, DOI 10.1007/s10826-013-9889-9.

⁶¹Suzuki, H., Bryant, F.B., & Edwards, J.D. (2010). Tracing prospective profiles of juvenile delinquency: An optimal classification tree analysis. *Optimal Data Analysis, 1*, 125-143.

⁶²Taft, C.T., Pless, A.P., Stalans, L.J. Koenen, K.C., King, L.A., & King, D.W. (2005). Risk factors for partner violence among a national sample of combat veterans. *Journal of Consulting and Clinical Psychology, 73*, 151-159.

⁶³Yarnold, P.R. (2013). Analyzing categorical attributes having many response categories. *Optimal Data Analysis, 2*, 172-176.

⁶⁴Yarnold, P.R. (2013). Assessing hold-out validity of CTA models using UniODA. *Optimal Data Analysis, 2*, 31-36.

⁶⁵Yarnold, P.R. (2013). Initial use of hierarchically optimal classification tree analysis in medical research. *Optimal Data Analysis, 2*, 7-18.

⁶⁶Yarnold, P.R. (2013). Univariate and multivariate analysis of categorical attributes with many response categories. *Optimal Data Analysis, 2*, 177-190.

⁶⁷Yarnold, P.R. (2014). "Predicting in-hospital mortality of patients with AIDS-related *Pneumocystis carinii* pneumonia: An example of hierarchically optimal classification tree analysis" (Yarnold et al., *Statistics in Medicine*, 1997, 16, 1451-1463): Corrected Illustration of CTA Model. *Optimal Data Analysis, 3*, 28-29.

⁶⁸Yarnold, P.R. (2014). Illustrating how 95% confidence intervals indicate model redundancy. *Optimal Data Analysis, 3*, 96-97.

⁶⁹Yarnold, P.R. (2014). Triage algorithm for chest radiography for community-acquired pneumonia of Emergency Department patients: Missing data cripples research. *Optimal Data Analysis, 3*, 102-106.

⁷⁰Yarnold PR, Soltysik RC (2010). Aggregated vs. referenced categorical attributes in UniODA and CTA. *Optimal Data Analysis, 1*, 46-49.

⁷¹Yarnold, P.R., & Soltysik, R.C. (2010). Manual vs. automated CTA: Optimal preadmission staging for inpatient mortality from *Pneumocystis carinii* pneumonia. *Optimal Data Analysis, 1*, 50-54.

⁷²Yarnold PR, Soltysik RC (2010). Maximizing the accuracy of classification trees by optimal pruning. *Optimal Data Analysis*, 1, 23-29.

⁷³Yarnold, P.R., & Soltysik, R.C. (2010). Unconstrained covariates in CTA. *Optimal Data Analysis*, 1, 38-40.

⁷⁴Yarnold, P.R., & Soltysik, R.C. (2013). Reverse CTA: an optimal analog to analysis of variance. *Optimal Data Analysis*, 2, 43-47.

⁷⁵Yarnold PR, Bryant FB, & Smith JH. (2013). Manual vs. automated CTA: Predicting freshman attrition. *Optimal Data Analysis*, 2, 48-53.

⁷⁶Yarnold, P.R., Michelson, E.A., Thompson, D.A., & Adams, S.L. (1998). Predicting patient satisfaction: a study of two emergency departments. *Journal of Behavioral Medicine*, 21, 545-563.

⁷⁷Yarnold, P.R., Soltysik, R.C., & Bennett, C.L. (1997). Predicting in-hospital mortality of patients with AIDS-related Pneumocystis carinii pneumonia: An example of hierarchically optimal classification tree analysis. *Statistics in Medicine*, 16, 1451-1463.

⁷⁸Yarnold, P.R., Soltysik, R.C., & Collinge, W. (2013). Modeling individual reactivity in serial designs: An example involving changes in weather and physical symptoms in fibromyalgia. *Optimal Data Analysis*, 2, 37-42.

⁷⁹Yarnold, P.R., & Soltysik, R.C. (2014). Globally optimal statistical classification models, I: Binary class variable, one ordered attribute. *Optimal Data Analysis*, 3, 55-77.

⁸⁰Yarnold, P.R., & Soltysik, R.C. *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books, 2005.

⁸¹Soltysik, R.C., & Yarnold, P.R. (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, 2, 194-197.

⁸²Soltysik, R.C., & Yarnold, P.R. (2013). MegaODA large sample and BIG DATA time trials: Harvesting the Wheat. *Optimal Data Analysis*, 2, 202-205.

⁸³Yarnold, P.R., & Soltysik, R.C. (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, 2, 220-221.

⁸⁴Yarnold, P.R. (1996). Discriminating geriatric and non-geriatric patients using functional status information: An example of classification tree analysis via UniODA. *Educational and Psychological Measurement*, 56, 656-667.

⁸⁵Soltysik, R.C., & Yarnold, P.R. (2010). Automated CTA software: Fundamental concepts and control commands. *Optimal Data Analysis*, 1, 144-160.

⁸⁶Soltysik, R.C., & Yarnold, P.R. (2013). Statistical power of optimal discrimination with one attribute and two classes: One-tailed hypotheses. *Optimal Data Analysis*, 2, 26-30.

⁸⁷Yarnold PR, Soltysik RC (2014). Globally optimal statistical classification models, II: Unrestricted class variable, two or more attributes. *Optimal Data Analysis*, 3, 78-84.

⁸⁸Yarnold, P.R., & Soltysik, R.C. (2014). Discrete 95% confidence intervals for ODA model- and chance-based classifications. *Optimal Data Analysis*, 3, 107-109.

Author Notes

Mail Address: Optimal Data Analysis LLC;
6348 N. Milwaukee Ave.; Chicago, IL 60646.