Milestones
Annual Report
2016

Stanley J. Hamstra, PhD
Laura Edgar, EdD, CAE
Kenji Yamazaki, PhD
Eric S. Holmboe, MD

-for the Milestones Team

Accreditation Council for Graduate Medical Education

401 North Michigan Avenue
Suite 2000
Chicago, Illinois 60611

Phone 312.755.5000
Fax 312.755.7498
www.acgme.org

October 2016

ACGME

You are a critical partner in the ACGME's mission to improve health care and population health. We do this together by advancing the quality of resident physicians' education and monitoring that quality through accreditation. The Milestones have been identified as a key component of this process, as they allow for continuous tracking of skills development and competency of resident/fellow physicians throughout their training. Until now, this was a conceptual promise, without data. The document presented here is the first national report of such data.

We wish to thank the program directors, as well as the members of the Clinical Competency Committees, faculty members and the rest of the GME community for developing and managing a complex system for teaching and collecting Milestones data. In this report, we present national Milestones data in aggregate form. This report also includes suggestions for how you might use the data to make improvements in your program.

While it is still too early to present individual program data in any meaningful way, we anticipate that our next Milestones Annual Report in 2017 will contain program-level data within each specialty. Please stay tuned as the necessary research to determine the best way to interpret and communicate those data is ongoing. In the meantime, we will continue to reach out to program director groups and other key stakeholders to determine the best strategies for interpretation of these data and how they might be used to improve curriculum and assessment processes.

Thank you for your support in this process and for continuing to collaborate with us to make this assessment process efficient, and to make the data truly reflective of your residents' and/or fellows' underlying competence.

Finally, please consult our online resources and tools for more information, including the *Milestones Guidebook*, Milestones FAQs, and the *Clinical Competency Committee Guidebook*, all available on the Milestones section of the ACGME website. We will continue to add resources to the website, so check back frequently.

Sincerely,
The Milestones Team

Executive Summary

The ACGME's mission is to improve health care and population health by advancing the quality of resident physicians' education through accreditation. Since 2012, the Milestones have been identified as a key component of this process, as they allow for continuous tracking of skills development and competency of resident and fellow physicians throughout their training.

The ACGME has now been collecting Milestones data on all resident and fellow physicians since 2015. This report is a snapshot of Milestones ratings from June 2016. It is the first annual report of Milestones data, intended to highlight both central tendencies and meaningful variation within specialties.

Overarching Themes

1) Across all specialties, the central tendency of the data show general attainment of the Milestones across years in program.
2) Each specialty shows variation in attainment of the Milestones across residents and programs, which needs to be investigated further.

How to Use These Findings

This report documents in detail the Milestones data across the various levels of Milestones achievement within each specialty. While the data are presented in aggregate form across programs and trainees, the report also provides some indication of the variation that occurs between trainees within a specialty. By presenting the data in this manner, the individual program director should be able to compare his/her local data with national trends within the program's specialty.

While it is still too early to present individual program-specific interpretive analyses, this report presents the case for the potential value of the data. Readers are encouraged to reflect on the data presented here to deepen understanding of exactly what they represent and how they might be useful in effecting change.

Intended Audience

The intended audience for this report includes program directors, leaders within specialty societies who oversee the development of national curricula, Review Committees that oversee accreditation of individual training programs, and the residents and fellows who are being trained within these programs. Future reports will be developed for other stakeholder groups, such as specialty boards, designated institutional officials (DIOs), policymakers, and the public.

Much of these data have already been shared with specialty societies, program director associations, and focus groups of program directors at educational and academic conferences. Together, the Milestones Team has presented relevant data to 19 of the 28 core specialties, in addition to nearly all of the Review Committees (in aggregate form).

Table of Contents

Background

Milestones in the Broader Context of Competency-Based Medical Education

In 2012 ACGME introduced the Next Accreditation System (NAS) for improving graduate medical education (GME).[1] This evolution in the accreditation model is intended to help address the changing health care needs of the population, to directly address the Institute for Healthcare Improvement (IHI) Triple Aim, and to improve educational and clinical outcomes. [1-3]

An important component of the NAS is a shift towards competency-based medical education. The use of the educational Milestones is seen as a way to facilitate the transformation from a process-bound system of GME accreditation to one that focuses on educational and clinical outcomes. An outcomes-based training system has the potential to better prepare physicians for a changing health care system by emphasizing a focus on the abilities of graduating residents and fellows and ensuring they match patient and health care system needs.[4-6]

The educational Milestones allow for continuous monitoring and quality improvement for GME. Specialties (and individual programs) can now focus on making improvements that align with the specific competencies outlined by the Milestones within their specialty, up to the point of graduation. This will help to ensure that graduates from these programs meet the expected standards of the profession and the goal of unsupervised practice.

Meeting the Challenge of Professional Self-Regulation

Tracking of Milestones data for individual residents and fellows is an essential part of the commitment of the ACGME to meet the responsibility of self-regulation,[1,3,7] and the expectations for quality and safety from its primary stakeholder, the public. It is important to note that the primary goal of the ACGME continues to be accreditation of institutions and residency and fellowship programs and ensuring the quality of those institutions and programs. The collection of individual residents' and fellows' Milestones data will only be used in aggregate form by the ACGME to address this goal. The ACGME will continue to collaborate with the American Board of Medical Specialties (ABMS) and the various specialty boards toward ensuring high quality graduates and maintaining the public trust.

Using Milestones Data to Provide Useful Feedback to Residency Programs

The spirit of the NAS implies an educational continuous quality improvement (CQI) framework. The collection, sharing, and interpretation of Milestones data allows for refinement of policies and accreditation standards for effective training. These data will provide the empirical basis for working with specialty groups and programs that may require guidance to meet the expectations set out in the NAS and the IHI Triple Aim. In essence, the field must work together to ensure residents and fellows are ready for unsupervised practice following graduation, and the Milestones can provide detailed guidance for achieving this goal.

Development of the Milestones

Individual Milestones were developed based on sub-competencies within each specialty.[8] The process of creating these detailed developmental markers of progression involved over 916 subject matter experts from 250 institutions. The Milestone Working Groups included members of the ACGME Review Committees, representatives from the ABMS member certification boards, program directors, residents and fellows, and representatives from specialty societies. These subject matter experts were guided by Advisory Groups from the ACGME, as well as a survey of program directors on content.

While most of the groups were very familiar with the traditional competencies of Medical Knowledge and Patient Care, there was a concerted attempt to meet the larger health care needs of the population and the IHI Triple Aim by expanding and specifying in detail competencies relating to Practice-Based Learning and Improvement (PBLI), Systems-Based Practice (SBP), Professionalism (PROF), and Interpersonal and Communication Skills (ICS).

Working Groups were instructed to develop the specific Milestones and establish a recommended graduation target using a developmental framework of skills acquisition, usually progressing explicitly from Level 1-5.[4-5] The Milestones were written largely independently by the 28 ACGME-accredited specialties and their associated subspecialties, covering the six Core Competencies, and resulting in an average of 24 sub-competencies per specialty (range 10-43), with training periods that range from single-year fellowships to residency programs lasting seven years. This complexity precludes the development of a single approach to data analysis and interpretation.

The Role of Key Stakeholders in Meaningful Interpretation of the Milestones Data

To make sense of such a large and complex dataset, these data are being presented first to a key stakeholder group, residency program directors and leaders of specialty societies, to assist in constructing meaningful interpretation with the goal of improving GME. By taking the perspective of the needs of key stakeholders, success in this endeavor will depend more on how these stakeholders *interpret* the data than how they are summarized and analyzed. In other words, the effectiveness of Milestones data in achieving the vision of the NAS depends as much on understanding the context in which the data are analyzed, and the dialogue with stakeholders regarding proper *interpretation* than it does on in-house strategies for analysis. This is consistent with recent advances in the field of *Implementation Science*.[9]

A Call to Action for Program Directors

Program directors and specialty societies represent the best source of information for understanding what these data represent. Without the rich contextual knowledge, we can only speculate and offer generalized interpretations based on theory or insights from other areas of application. It would be most helpful if program directors considered in detail what might explain the observed patterns in the data reported in Tables 3-31 and share this with their respective wider community, including other program directors,

leaders in the relevant specialty society, and the ACGME Milestones Team in order to ensure the data are interpreted properly and ultimately fulfill their potential to improve GME.

How to Use These Findings

Two particularly important items to highlight in this first Milestones Annual Report relate to guiding changes in curriculum and the development of better assessment methods. As can be seen in the data presented in Tables 3-31, there are competency areas where the range of Milestone attainment is wide. Gaps in curriculum and effective assessment are significant contributors to the observed variation in these early findings. This should be viewed as welcome news as it is an early signal the Milestones data are providing useful information to drive continuous quality improvement (CQI) in GME. The importance of these purposes and goals in these early phases of the Milestones rollout cannot be overstated.

Constituency or Stakeholder	**Purpose/Function**
Residents and Fellows	• Provide a descriptive roadmap for training • Increase transparency of performance requirements • Encourage informed self-assessment and self-directed learning • Facilitate better feedback to trainee
Residency and Fellowship Programs	• Guide curriculum and assessment tool development • Provide meaningful framework for the Clinical Competency Committee (CCC) (e.g., help create shared mental model) • Provide more explicit expectations of residents and fellows • Enhance opportunity for early identification of under-performers
ACGME and the Public	• Public Accountability – report at an aggregated national level on competency outcomes • Build community for evaluation and research, with focus on continuous quality improvement
Certification Boards	• Enable ongoing research to improve certification processes

Methods

Every six months, the ACGME receives over 2.8 million data points based on performance assessments of over 120,000 learners from residency and fellowship programs across the US. The sheer size and complexity of this dataset make it impossible to provide general statements about resident or fellow progression and suggestions for educational CQI for individual residency or fellowship training programs.

Ongoing Development of Strategies for Interpretation

The Department of Milestones Research and Evaluation regularly consults with advisory groups composed of leading experts in medical education from across the country to help us develop strategies for analysis and interpretation of the data and the processes that lead to Milestones ratings and reporting.

It is useful here to consider that a resident's or fellow's competence in a particular aspect of clinical practice is the target construct that should be represented in the Milestones data. Thus, to interpret Milestones data correctly, there must be assurances that the data accurately represent a resident's or fellow's competence, instead of other variables that might influence the Milestone ratings, including curriculum factors, quality of assessment tools, and the ability to observe the resident or fellow in the variety of clinical settings that make up that individual's training experience in a complex clinical environment. All of these factors can affect current Milestone judgments. Examining the data in Tables 3-31 in terms of "learning curves" of gradual progression of competence shows that:

- there is variation in Milestone attainment across specialties and competencies; and,
- not all residents/fellows in all programs reach Level 4 in all competencies by graduation.

There are likely several possible reasons for this, including:

- differences in the complexity of the Milestones competency language as originally written for that specialty;
- differences in clinical exposure of some residents in some programs;
- variation in scoring by raters;
- differences in the quality of assessment rating forms; or,
- differences in the types of assessment methods used to show attainment of the Milestones.

Strategies for Communication and Implications for Change

Of course, many other factors may be at play, and these are the subject of intensive, ongoing research. Until this research is mature, the data should only be reported in the context of interpretive statements and assumptions that are relevant to the particular stakeholder group, (i.e., DIOs, program directors, residents/fellows, the public). In

communicating the results of Milestones analysis, there must be full awareness of the consequences of the analysis and communication, which implies diligence in providing context and guidance for interpretation when presenting these results.[10,11]

Overarching Themes

1) Nearly all programs are reporting Milestones data every six months to the ACGME (for the latest period, the reporting rate was 99.995%).

2) Generally, across all specialties and all programs, the Milestones data show attainment of Milestones across years in program (See Tables 3-31).

3) Each specialty shows variation in attainment of the Milestones, which needs to be investigated further (See Tables 3-31).

Further insights gained from numerous presentations, focus groups, and interactions with key stakeholders across the country include the following:

General Recommendations

- It is best not to use the Milestones themselves as assessment tools for residents/fellows on rotations or short clinical experiences.
- It is generally better to have a comprehensive "system" (program) of assessment rather than an *ad hoc* collection of disparate assessment tools.
- It appears to be most effective if residents/fellows are engaged in the collection of their performance data so they can more readily respond to areas for improvement.
- The process needs to be reasonable and easy to do.
- There needs to be further development in demonstrating exactly how to achieve certain milestones.
- There is a desire for improved assessment tools for skills related to ICS and PROF.

Early Benefits Reported

- There were many reports of more structure now in the program, highlighting a more comprehensive approach to monitoring residents'/fellows' progression throughout their training.
- Many residents/fellows reported receiving more comprehensive and structured feedback.
- Milestones more easily highlight specific gaps in training.
- Both residents/fellows and faculty members are more comfortable now with the narrative aspect of the Milestones vs. the "numbers."

Early Challenges Reported

- The process can be onerous and make for considerable extra work for some programs.
- The Milestones as written often do not reflect the underlying construct.

- In some cases, residents/fellows fall between the levels specified on the ACGME Milestone Evaluations form.
- In some cases, there are too many sub-competencies, which don't seem to add much to the assessment process.

Future Directions

In conclusion, the Milestones data are currently very complex and caution must be exercised in how these results are interpreted and communicated to various stakeholders. The validity of the data is only beginning to emerge, and there are potentially serious implications for misinterpretation. A validity framework (currently under development) can guide the process of CQI and help to realize the vision of NAS as articulated by Nasca et al. in 2012.[1]

In response to the information received to date, a system-wide project has been implemented to ultimately revise the Milestones language to make it easier for program directors to understand and implement locally, as well as to examine areas in which the Milestones language can be harmonized across specialties, especially in Competencies such as PROF and ICS. This project has come to be known as "Milestones 2.0," and will take several years to complete, as feedback from the various stakeholders is collected.

In addition to work on Milestones 2.0, an obvious next step is to continue the interpretive work and research suggested above. This includes work on predictive validity of the Milestones, for example by correlating Patient Care Milestones ratings with independent measures of clinical performance or patient outcomes, as well as many other projects. To this end, several projects have been initiated by the Milestones Team, as well as collaboratively with academic colleagues, to help address the following specific areas:

1) ongoing work to revise the Milestones language within each specialty, but also to consider harmonizing the language for PROF, ICS, SBP, and PBLI across specialties;
2) ongoing research on predictive validity of Milestones ratings vs. Board scores in various specialties and other clinically-relevant performance data; and,
3) ongoing research on CCC processes to examine best practices for ensuring these committees take the necessary steps to maximize the validity of the Milestones ratings they report to the ACGME.

This work is ongoing and will continue to appear in the peer-reviewed literature to help build a stronger evidence base for the ability of this accreditation model to meet the larger vision outlined in the Introduction above. Finally, every member of the GME community should engage in research and debate regarding the potential for Milestones data to effect meaningful change in GME.

Limitations

While the interpretations and conclusions that could be drawn from the data presented in this report are based on a single point in time (i.e., June 2016), trends for stability in the data patterns for academic year-end since June 2014 have recently been examined for the Phase 1 reporting specialties. Most of the Milestones data show signs of stability across this period, which lends greater confidence to the potential interpretations and conclusions that can be drawn from them, and will allow for greater confidence in communication of these interpretations back to the community to complete the CQI loop.

Acknowledgments

The authors wish to thank the program directors who attended the feedback sessions at both the 2015 and the 2016 ACGME Annual Educational Conference for their valuable insights, as well as the various stakeholder groups to whom preliminary versions of these results have been presented over the past year.

Contact Us:

milestones@acgme.org

or:

Stan Hamstra, VP, Milestones Research and Evaluation:
shamstra@acgme.org

Eric Holmboe, Senior VP, Milestone Development and Evaluation:
eholmboe@acgme.org

Laura Edgar, Executive Director, Milestones Development:
ledgar@acgme.org

References

1. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next accreditation system: rationale and benefits. *N Engl J Med*. 2012:366;1051–1056.
2. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Affairs* (Millwood). 2008 May-Jun;27(3):759-69.
3. IOM (Institute of Medicine). 2014. *Graduate medical education that meets the nation's health needs.* Washington, DC: The National Academies Press.
4. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. An antidote to over specification in the education of medical specialists. *Health Affairs*. 2002; 21: 103-111.
5. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. *Med Educ*. 2007; 41: 1124-30.
6. Hodges BD. "A Tea-Steeping or i-Doc Model for Medical Education?," Acad *Med* 85(9) Sept Suppl 2010, pp. S34-S44.
7. Nasca TJ. Professionalism and its implications for governance and accountability of graduate medical education in the United States. *JAMA*. 2015;313(18):1801-2.
8. Sullivan G, Simpson D, Cooney T, Beresin E. A Milestone in the Milestones movement: the JGME Milestones supplement. *J Grad Med Educ*. March, 2013 (suppl.): 1-4.
9. McGaghie WC. Implementation science. Addressing complexity in medical education. *Med Teach*. 2011; 33: 97-8.
10. Messick S. Validity. In: Linn RL, ed. *Educational Measurement*. 3rd Ed. New York, NY: American Council on Education and Macmillan; 1989:13–103.
11. Hubley AM, Zumbo BD. Validity and the consequences of test interpretation and use. *Soc Indic Res*. 2011;103:219-230.

Additional Resources

See Appendix A for a list of recent publications involving the Milestones. In addition, the ACGME website contains the *Milestones Guidebook*, the Milestones FAQs, and the *Clinical Competency Committee Guidebook*, as well as a copy of a recent descriptive paper by the ACGME Milestones Team entitled "Reflections on the First 2 Years of Milestone Implementation."

Visit https://www.acgme.org/What-We-Do/Accreditation/Milestones/Overview and scroll down to see the "Milestones Publications" menu in the lower right-hand side of the page.

Forthcoming papers/papers under review include:

- a study on perceptions of Milestones by faculty members and residents in neurological surgery, along with insights about best practices for CCC processes
- a study correlating Medical Knowledge Milestones ratings with Board exam scores in emergency medicine
- a study examining the impact of a simple intervention on the validity of Milestones ratings in emergency medicine
- a study comparing end of training evaluations with and without Milestones vs. Board exam scores in internal medicine
- a conceptual paper outlining a validity framework to help in interpreting Milestones ratings

Table 1 – Number of Sub-Competencies by Specialty

	Number of Sub-Competencies						
Specialty Name	**Total**	**PC**	**MK**	**SBP**	**PBLI**	**PROF**	**ICS**
Allergy and Immunology	10	4	1	1	2	1	1
Anesthesiology	25	10	1	2	4	5	3
Colon and Rectal Surgery	21	8	9	1	1	1	1
Dermatology	28	7	5	4	4	3	5
Emergency Medicine	23	14	1	3	1	2	2
Family Medicine	22	5	2	4	3	4	4
Internal Medicine	22	5	2	4	4	4	3
Internal Medicine/Pediatrics	43	10	3	7	8	10	5
Medical Genetics and Genomics	20	9	2	2	2	3	2
Neurological Surgery	24	8	8	2	2	2	2
Neurology	29	18	3	2	2	2	2
Nuclear Medicine	19	5	7	2	2	1	2
Obstetrics and Gynecology	28	11	7	2	2	3	3
Ophthalmology	24	8	2	3	3	4	4
Orthopaedic Surgery	41	16	16	3	2	2	2
Otolaryngology	17	8	4	2	1	1	1
Pathology	27	7	3	7	2	6	2
Pediatrics	21	5	1	3	4	6	2
Physical Medicine and Rehabilitation	19	7	1	3	3	3	2
Plastic Surgery	36	14	14	3	2	2	1
Plastic Surgery-Integrated	36	14	14	3	2	2	1
Preventive Medicine-Aerospace Medicine	27	15	4	3	1	2	2
Preventive Medicine-Occupational Medicine	26	14	4	3	1	2	2
Preventive Medicine-Public Health and General Preventive Medicine	23	12	3	3	1	2	2
Psychiatry	22	5	6	4	3	2	2
Radiation Oncology	22	11	2	3	2	2	2
Radiology-Diagnostic	12	2	2	2	3	1	2
Surgery	16	3	2	2	3	3	3
Thoracic Surgery-Integrated	26	8	10	3	2	2	1
Transitional Year	23	7	2	3	3	4	4
Urology	32	9	1	4	7	6	5
Vascular Surgery-Integrated	31	8	12	3	3	3	2

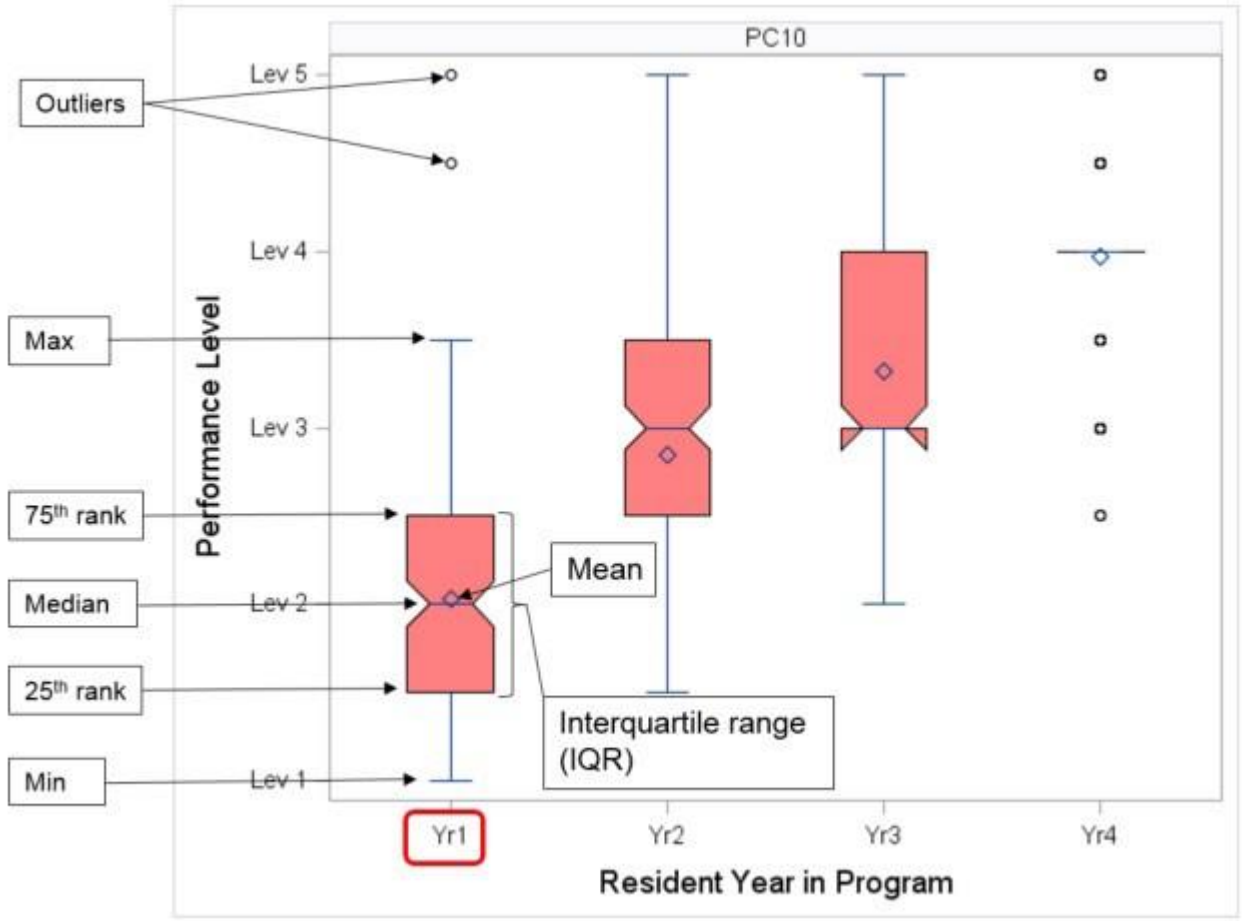
Note: **PC - Patient Care** **PBLI - Practice-Based Learning and Improvement**
MK - Medical Knowledge **PROF - Professionalism**
SBP - Systems-Based Practice **ICS - Interpersonal and Communication Skills**

Table 2 – Number of Residents by Year in Program

	Number of Residents - June 2016							
Specialty Name	**Total**	**Yr 1**	**Yr 2**	**Yr 3**	**Yr 4**	**Yr 5**	**Yr 6**	**Yr 7**
Allergy and Immunology	293	144	149					
Anesthesiology	5283	1082	1425	1398	1378			
Colon and Rectal Surgery	94	94						
Dermatology	1309	460	434	415				
Emergency Medicine	4290	1458	1420	1396	16			
Family Medicine	10612	3656	3549	3407				
Internal Medicine	25326	10117	7784	7425				
Internal Medicine/Pediatrics	1465	380	369	356	360			
Medical Genetics and Genomics	65	33	32					
Neurological Surgery	1402	229	223	218	202	190	199	141
Neurology	1457	390	381	379	307			
Nuclear Medicine	87	42	23	22				
Obstetrics and Gynecology	5238	1327	1311	1303	1297			
Ophthalmology	1439	487	480	472				
Orthopaedic Surgery	3892	790	791	790	764	757		
Otolaryngology	1563	320	320	314	307	302		
Pathology	2331	640	614	587	490			
Pediatrics	8952	3056	3028	2868				
Physical Medicine and Rehabilitation	743	251	247	245				
Plastic Surgery	364	104	128	132				
Plastic Surgery-Integrated	664	147	134	125	105	81	72	
Preventive Medicine-Aerospace Medicine	40	31	9					
Preventive Medicine-Occupational Medicine	131	61	70					
Preventive Medicine-Public Health and General Preventive Medicine	164	93	71					
Psychiatry	5409	1468	1474	1431	1036			
Radiation Oncology	737	192	190	173	182			
Radiology-Diagnostic	4763	1230	1193	1155	1185			
Surgery	8224	2717	1606	1367	1284	1250		
Thoracic Surgery-Integrated	139	35	33	30	19	13	9	
Transitional Year	1103	1103						
Urology	1264	327	312	319	306			
Vascular Surgery-Integrated	235	57	55	49	42	32		

Figure 1: Key to Box Plots

Box plots provide a rigorous and robust way to display complex data such as Milestones. The components of the box plots used for the Milestones are shown below. In this example, the data represent attainment of Patient Care sub-competency #10 (PC10).



As can be seen from this diagram, the **median** Milestone level for each year of resident is represented by the horizontal line, bounded by the **25th and 75th rank** of Milestone ratings, also known as the **inter-quartile range (IQR).** The **mean** rating is represented by the diamond, but should be interpreted with caution given Milestones are ordinal, not dimensional data. **Min** represents the lowest level and **Max** the highest level (the "**whiskers**"), excluding **outliers** (represented by the open circles). Overall we can see a general upward trajectory in this sub-competency from Year 1 (median level 2) to Year 4 (median level 4).

In this example, let's focus on Year 1 in the above box plot (highlighted by the red box). Remember that most Milestone sets possess five levels of development with transition zones between each Milestone level (designated as half increments such as level 2.5). Let's assume there are 2,000 Year 1 residents in this specialty. The median for Year 1 is level 2. The interquartile range is level 1.5 (25th percentile rank) and level 2.5 (75th percentile rank).

With regards to Milestone levels, we can sort the levels from least to greatest and then graph the levels as shown in this box-and-whisker plot. Take the highest 50% of the group (1,000) who were at or above Milestone level 2; they are represented by everything above the **median** line. 50% of the Year 1 residents fall between level 1.5 and level 2.5 (IQR). Those in the top 25% of Milestone judgments in the Year 1 group (500) are shown by the top "whisker" (here labeled as **Max**) and the **outlier** open circles. The outliers represent those who were judged to be substantially higher (in this case we see two outlier circles) or were judged to be a lot lower than normal (in these example there are no low outliers). The number of people represented by the circles will vary by the sample size.

Box-and-whisker plots also provide information on more than just the four split groups. You can also see which way the Milestone data can "sway". For example, if more residents are judged much higher than just a few residents being judged much lower, the median is going to be higher or the top whisker could be longer than the bottom one. Box-and-whisker plots provide a better overview of the Milestone data's distribution.

The Box-and-Whisker plots must be interpreted in the context of the Milestone descriptions for each sub-competency within each discipline. Provided below are links to each specialty Milestone set to help guide your review of the data.

Specialty:

Allergy and Immunology:
http://www.acgme.org/Portals/0/PDFs/Milestones/AllergyandImmunologyMilestones.pdf

Anesthesia: http://www.acgme.org/Portals/0/PDFs/Milestones/AnesthesiologyMilestones.pdf

Colon and Rectal Surgery:
http://www.acgme.org/Portals/0/PDFs/Milestones/ColonandRectalSurgeryMilestones.pdf

Dermatology: http://www.acgme.org/Portals/0/PDFs/Milestones/DermatologyMilestones.pdf

Emergency Medicine:
http://www.acgme.org/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf

Family Medicine: http://www.acgme.org/Portals/0/PDFs/Milestones/FamilyMedicineMilestones.pdf

Internal Medicine:
http://www.acgme.org/Portals/0/PDFs/Milestones/InternalMedicineMilestones.pdf

Medical Genetics and Genomics:
http://www.acgme.org/Portals/0/PDFs/Milestones/MedicalGeneticsMilestones.pdf

Neurological Surgery:
http://www.acgme.org/Portals/0/PDFs/Milestones/NeurologicalSurgeryMilestones.pdf

Neurology: http://www.acgme.org/Portals/0/PDFs/Milestones/NeurologyMilestones.pdf

Nuclear Medicine:
http://www.acgme.org/Portals/0/PDFs/Milestones/NuclearMedicineMilestones.pdf

Obstetrics and Gynecology:
http://www.acgme.org/Portals/0/PDFs/Milestones/ObstetricsandGynecologyMilestones.pdf

Ophthalmology: http://www.acgme.org/Portals/0/PDFs/Milestones/OphthalmologyMilestones.pdf

Orthopaedic Surgery:
http://www.acgme.org/Portals/0/PDFs/Milestones/OrthopaedicSurgeryMilestones.pdf

Otolaryngology: http://www.acgme.org/Portals/0/PDFs/Milestones/OtolaryngologyMilestones.pdf

Pathology: http://www.acgme.org/Portals/0/PDFs/Milestones/PathologyMilestones.pdf

Pediatrics: http://www.acgme.org/Portals/0/PDFs/Milestones/PediatricsMilestones.pdf

Physical Medicine and Rehabilitation:
http://www.acgme.org/Portals/0/PDFs/Milestones/PMRMilestones.pdf

Plastic Surgery: http://www.acgme.org/Portals/0/PDFs/Milestones/PlasticSurgeryMilestones.pdf

Preventive Medicine-Aerospace Medicine:
http://www.acgme.org/Portals/0/PDFs/Milestones/PreventiveMedicineMilestones-AerospaceMedicine.pdf

Preventive Medicine-Occupational Medicine:
http://www.acgme.org/Portals/0/PDFs/Milestones/PreventiveMedicineMilestones-OccupationalMedicine.pdf

Preventive Medicine-Public Health/General Preventive Medicine:
http://www.acgme.org/Portals/0/PDFs/Milestones/PreventiveMedicineMilestones-PublicHealthandGeneralPreventiveMedicine.pdf

Psychiatry: http://www.acgme.org/Portals/0/PDFs/Milestones/PsychiatryMilestones.pdf

Radiation Oncology:
http://www.acgme.org/Portals/0/PDFs/Milestones/RadiationOncologyMilestones.pdf

Radiology-Diagnostic:
http://www.acgme.org/Portals/0/PDFs/Milestones/DiagnosticRadiologyMilestones.pdf

Surgery: http://www.acgme.org/Portals/0/PDFs/Milestones/SurgeryMilestones.pdf

Thoracic Surgery:
http://www.acgme.org/Portals/0/PDFs/Milestones/ThoracicSurgeryMilestones.pdf

Transitional Year:
http://www.acgme.org/Portals/0/PDFs/Milestones/TransitionalYearMilestones.pdf

Urology: http://www.acgme.org/Portals/0/PDFs/Milestones/UrologyMilestones.pdf

Vascular Surgery:
http://www.acgme.org/Portals/0/PDFs/Milestones/VascularSurgeryMilestones.pdf

Table 3 - Allergy and Immunology (June 2016)

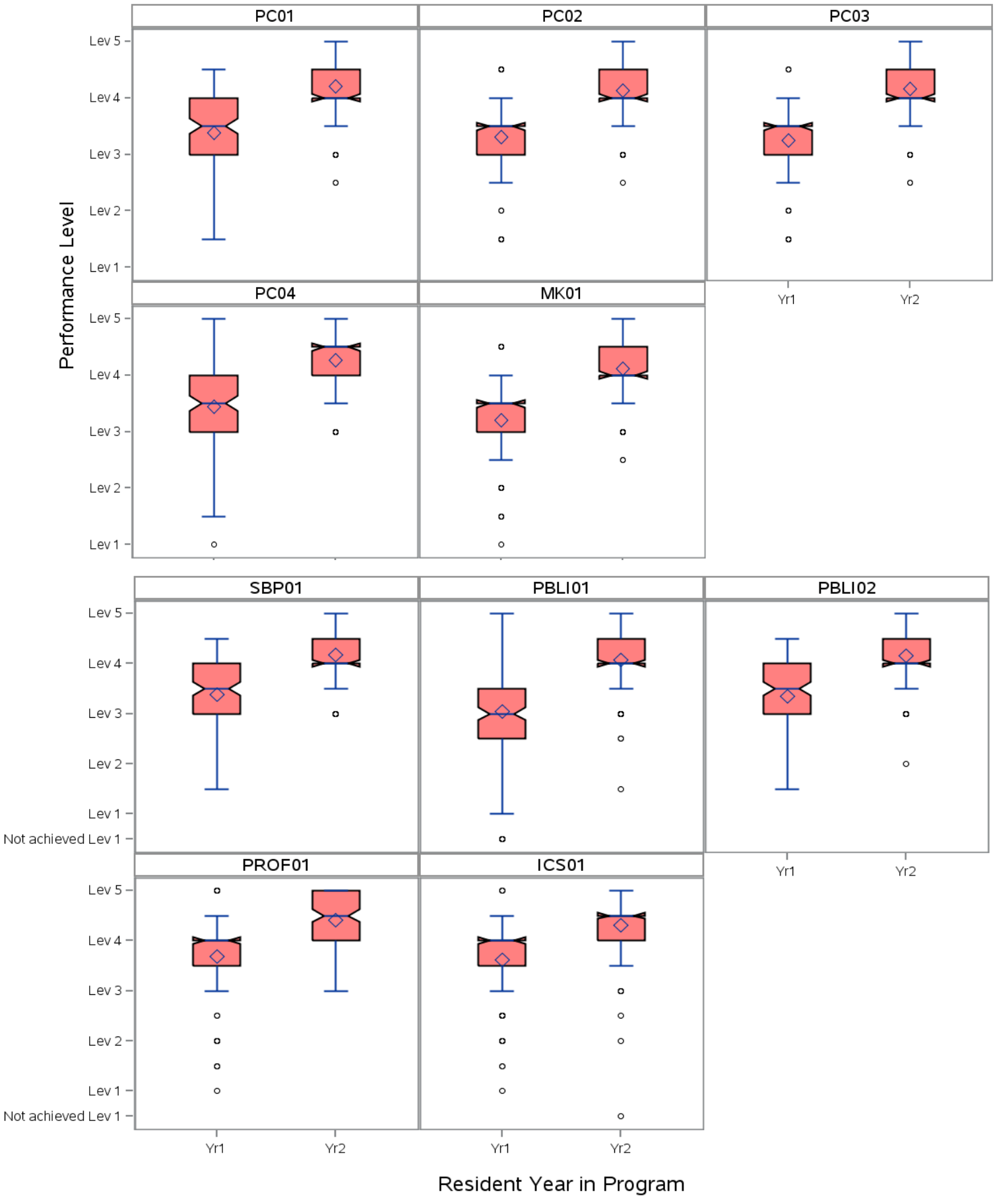


Table 4 - Anesthesiology (June 2016)

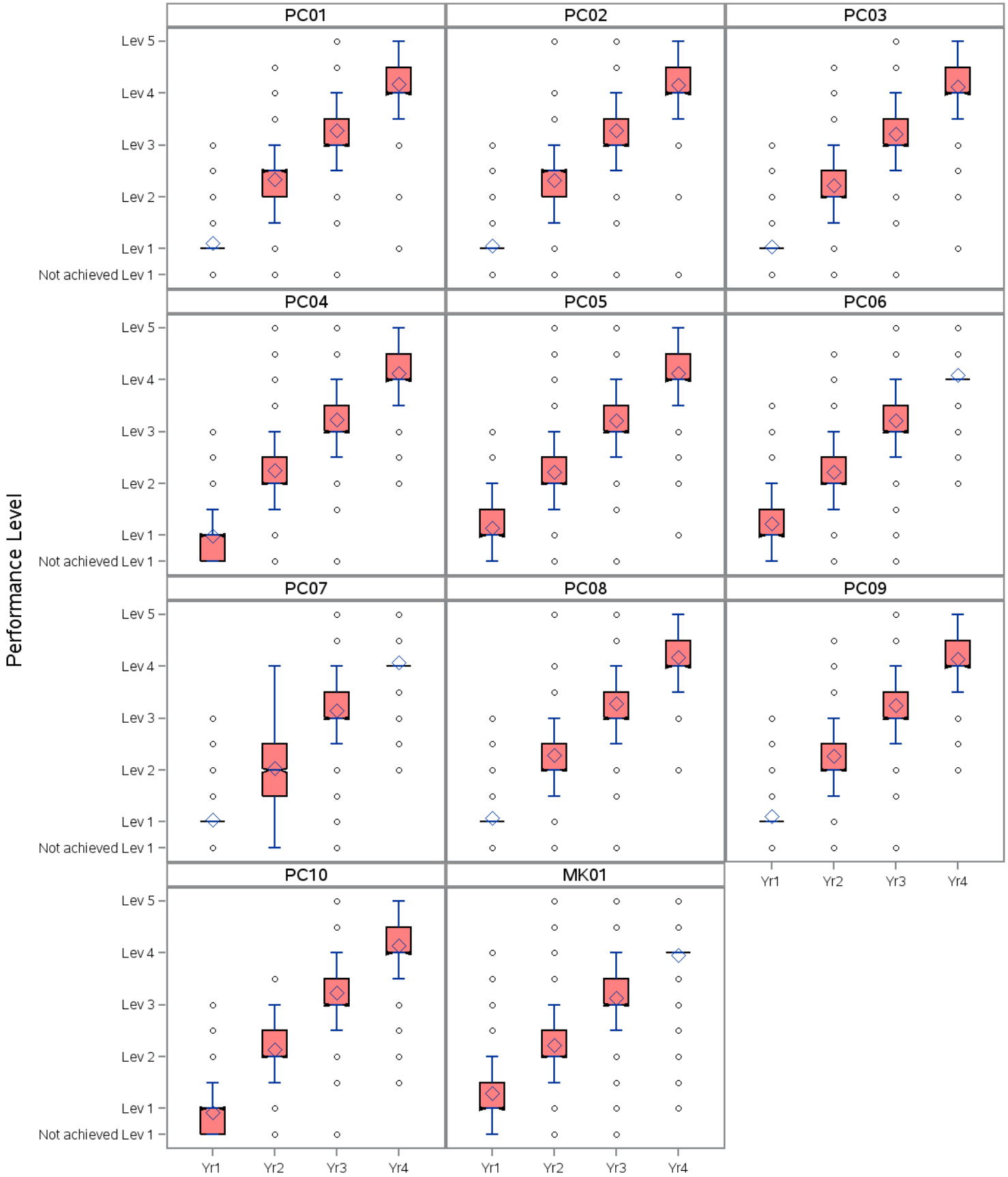


Table 4 - Anesthesiology (June 2016)

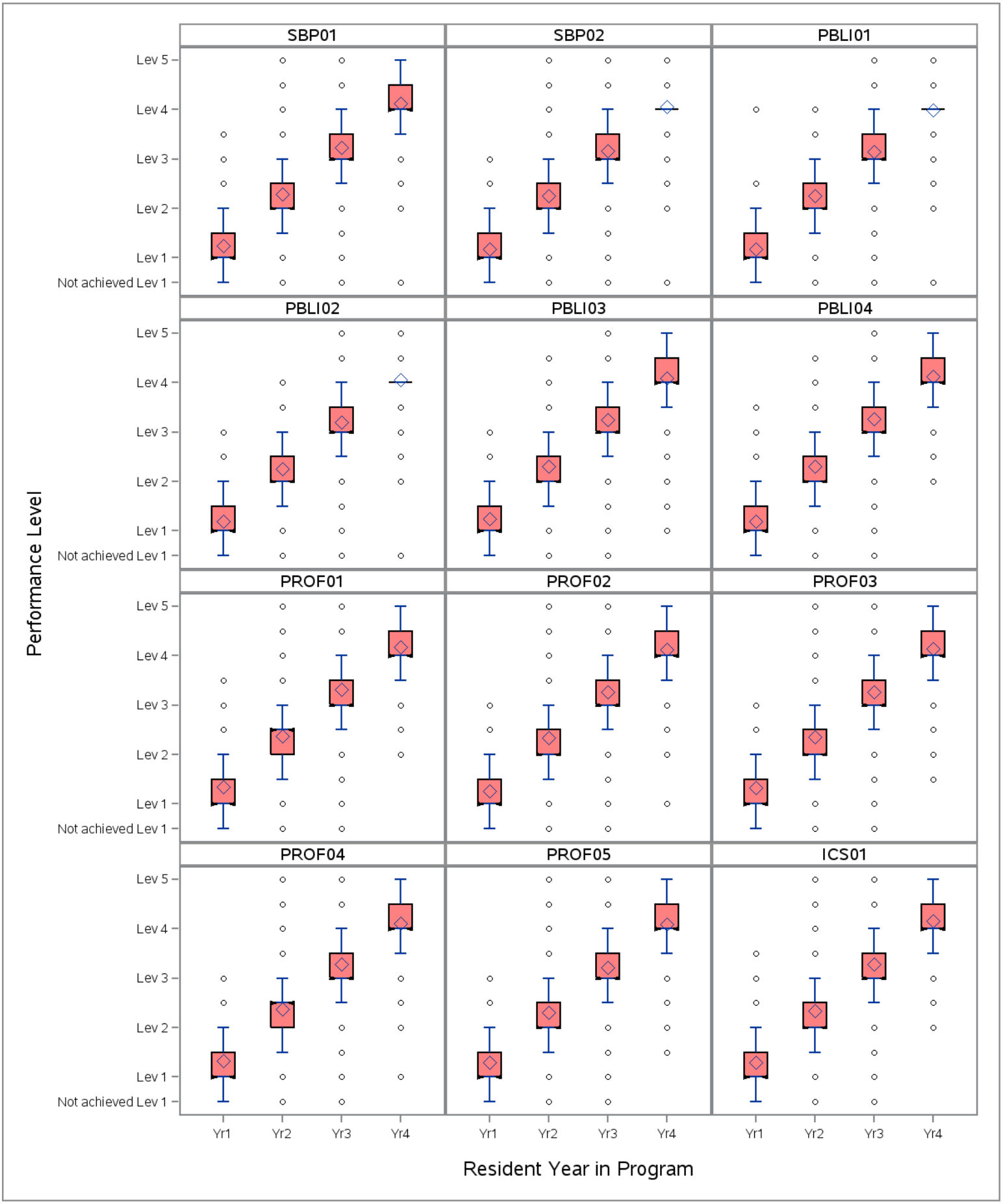


Table 4 - Anesthesiology (June 2016)

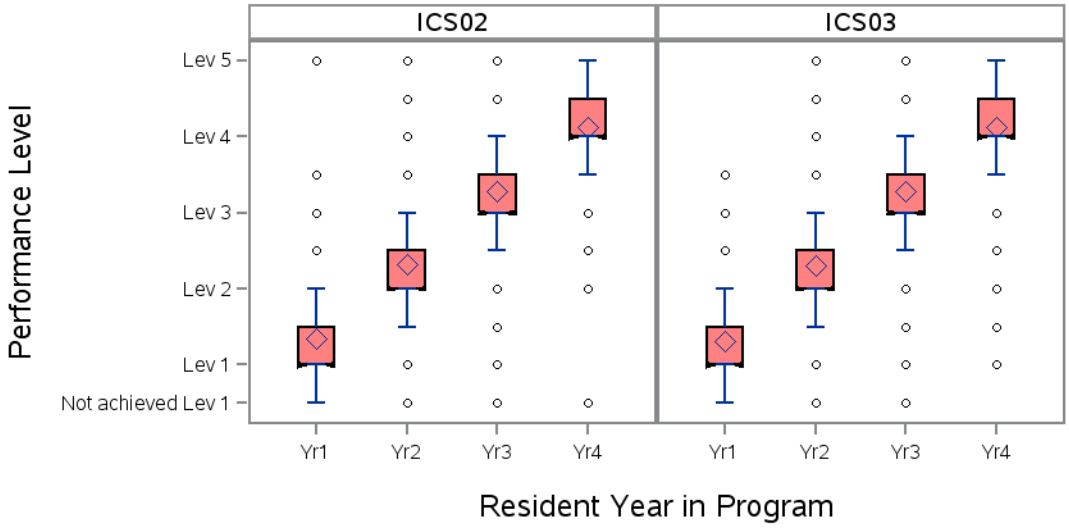


Table 5 - Colon and Rectal Surgery (June 2016)

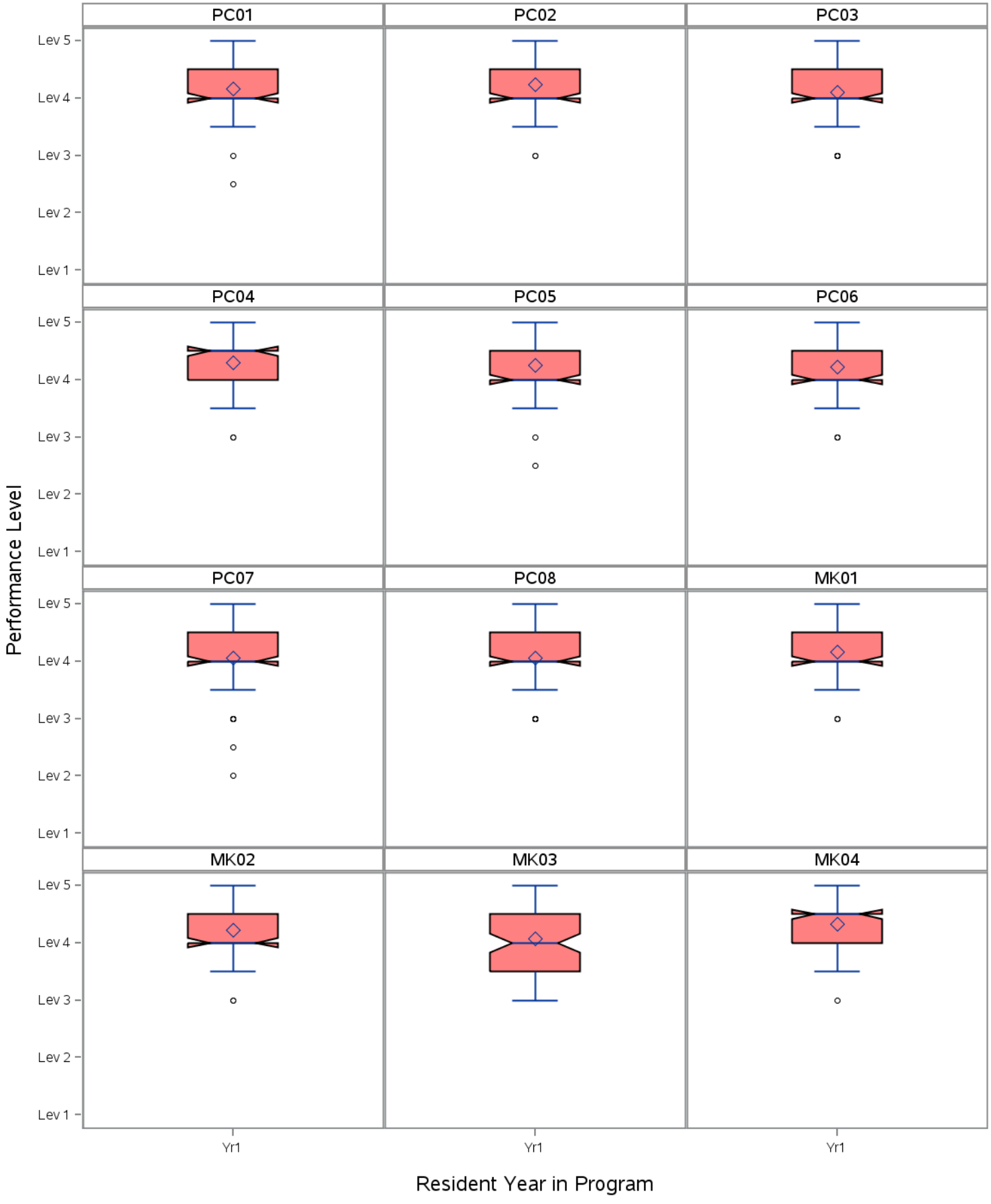


Table 5 - Colon and Rectal Surgery (June 2016)

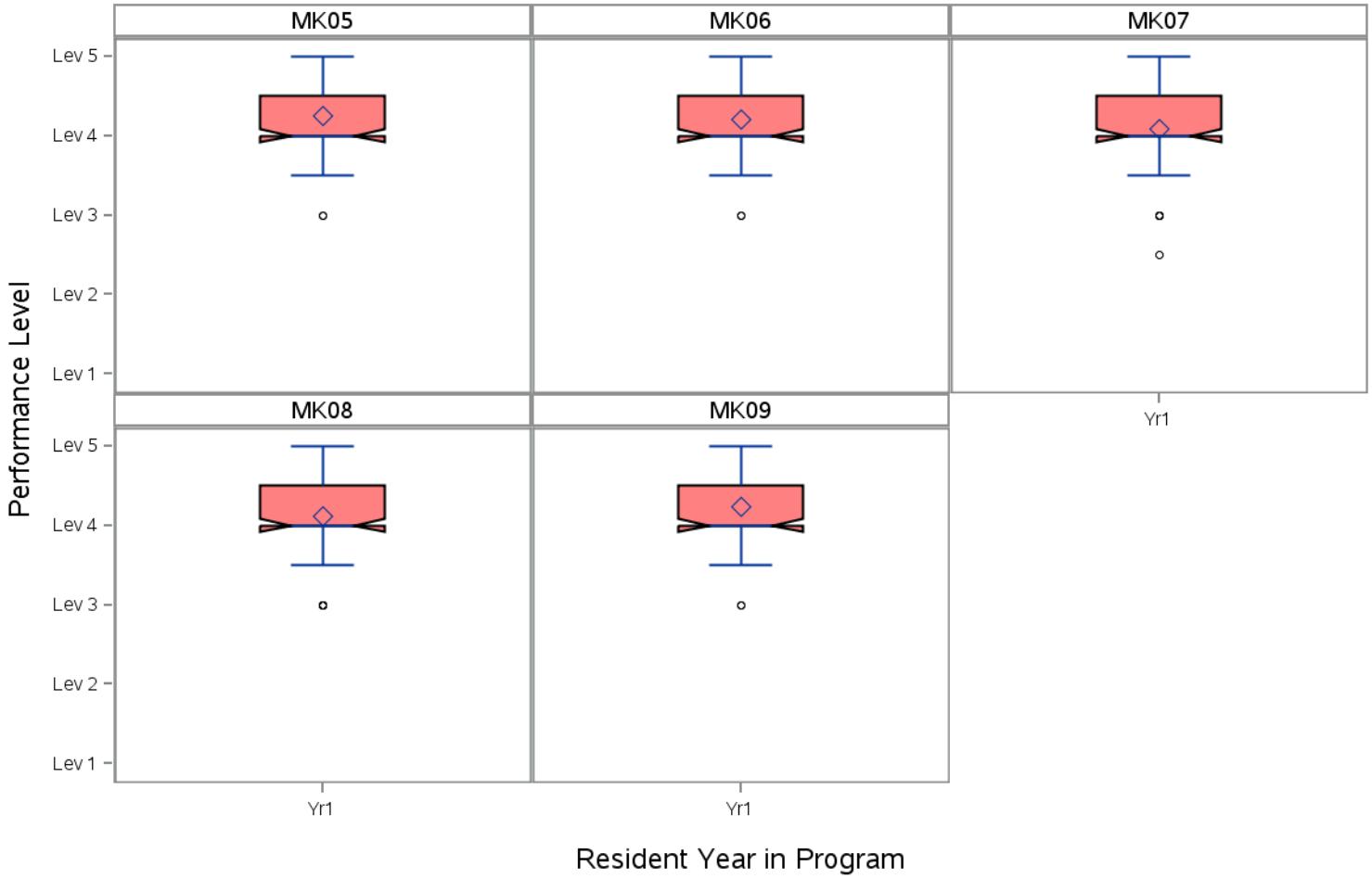


Table 5 - Colon and Rectal Surgery (June 2016)

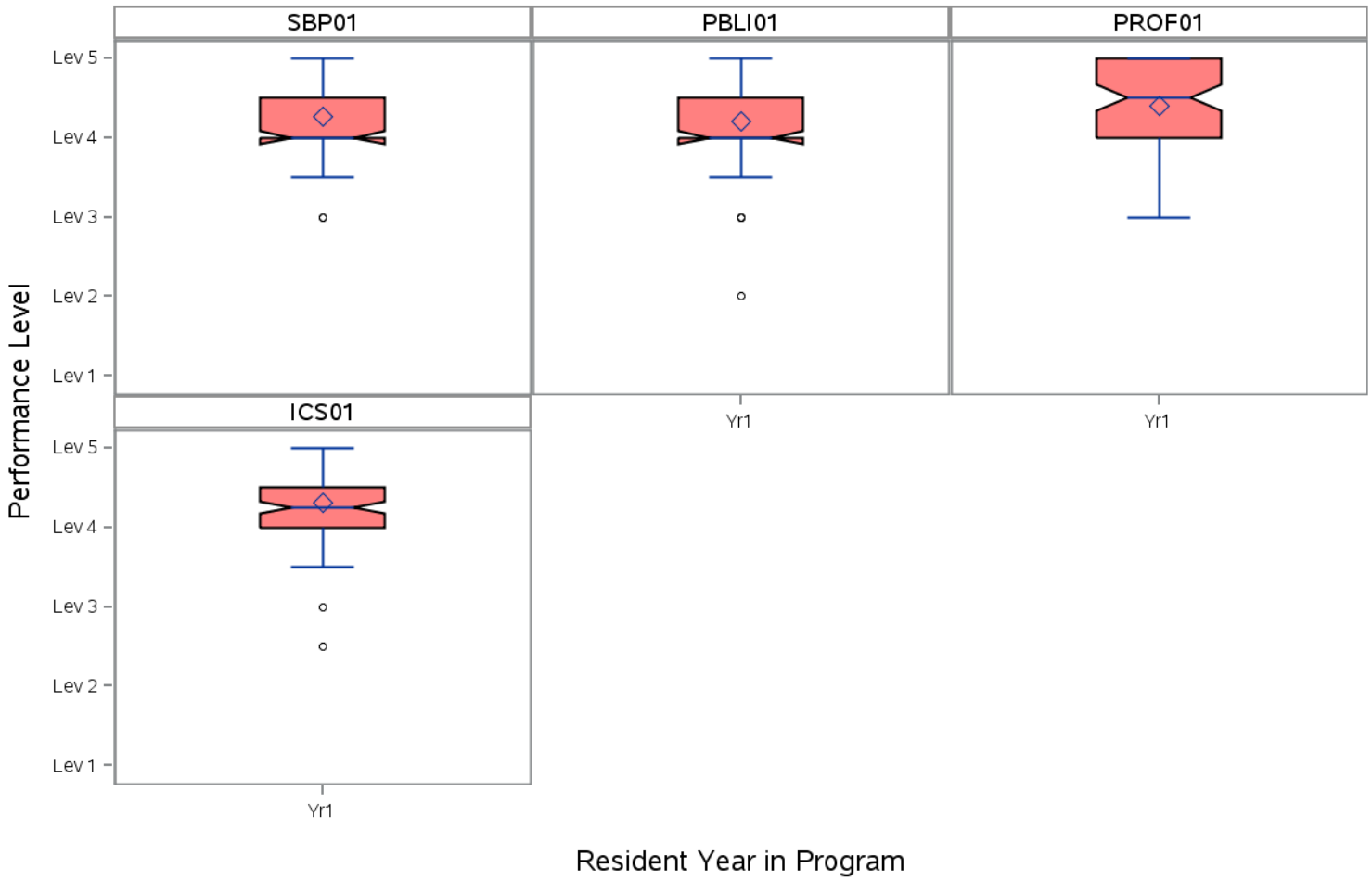


Table 6 - Dermatology (June 2016)

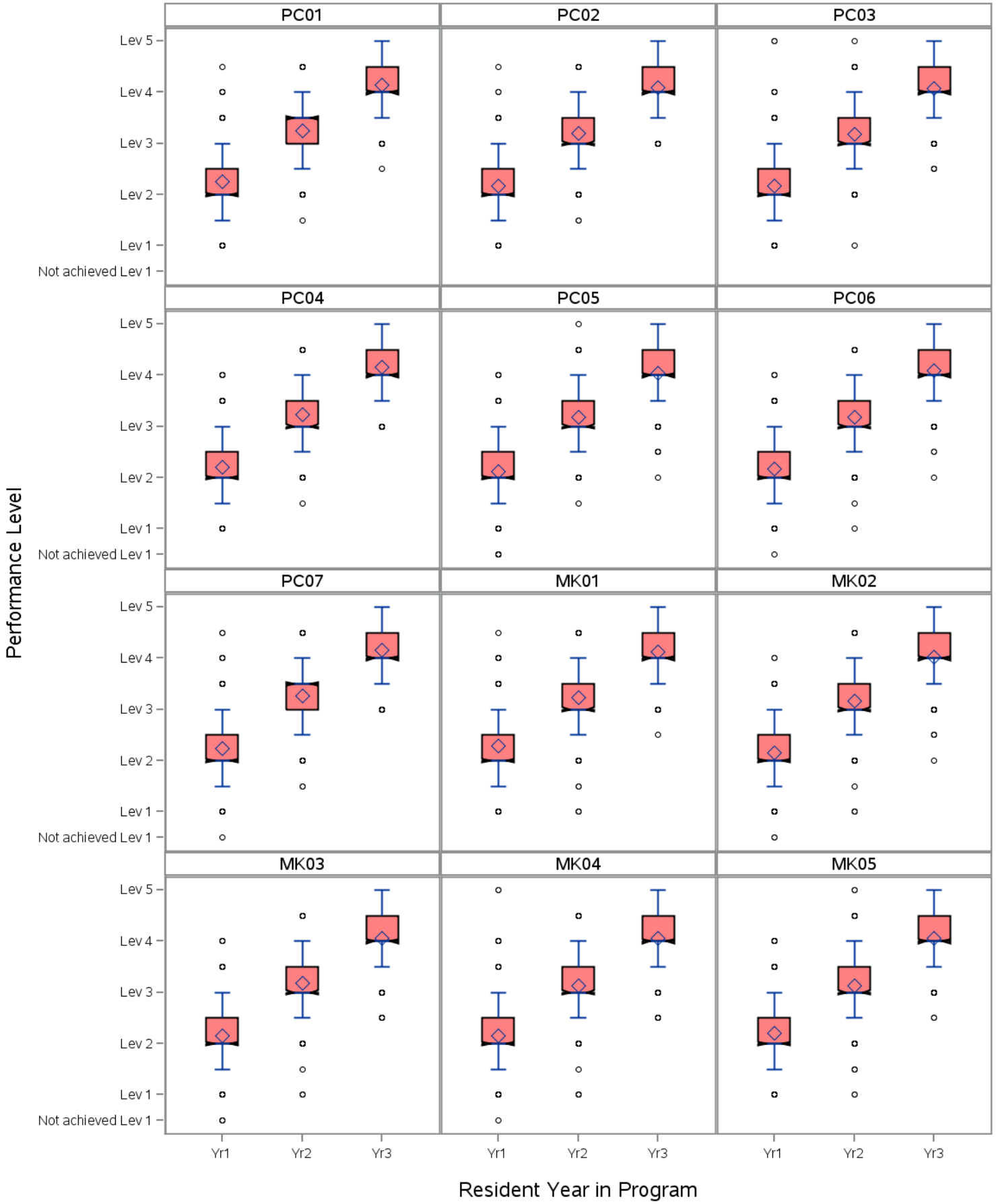


Table 6 - Dermatology (June 2016)

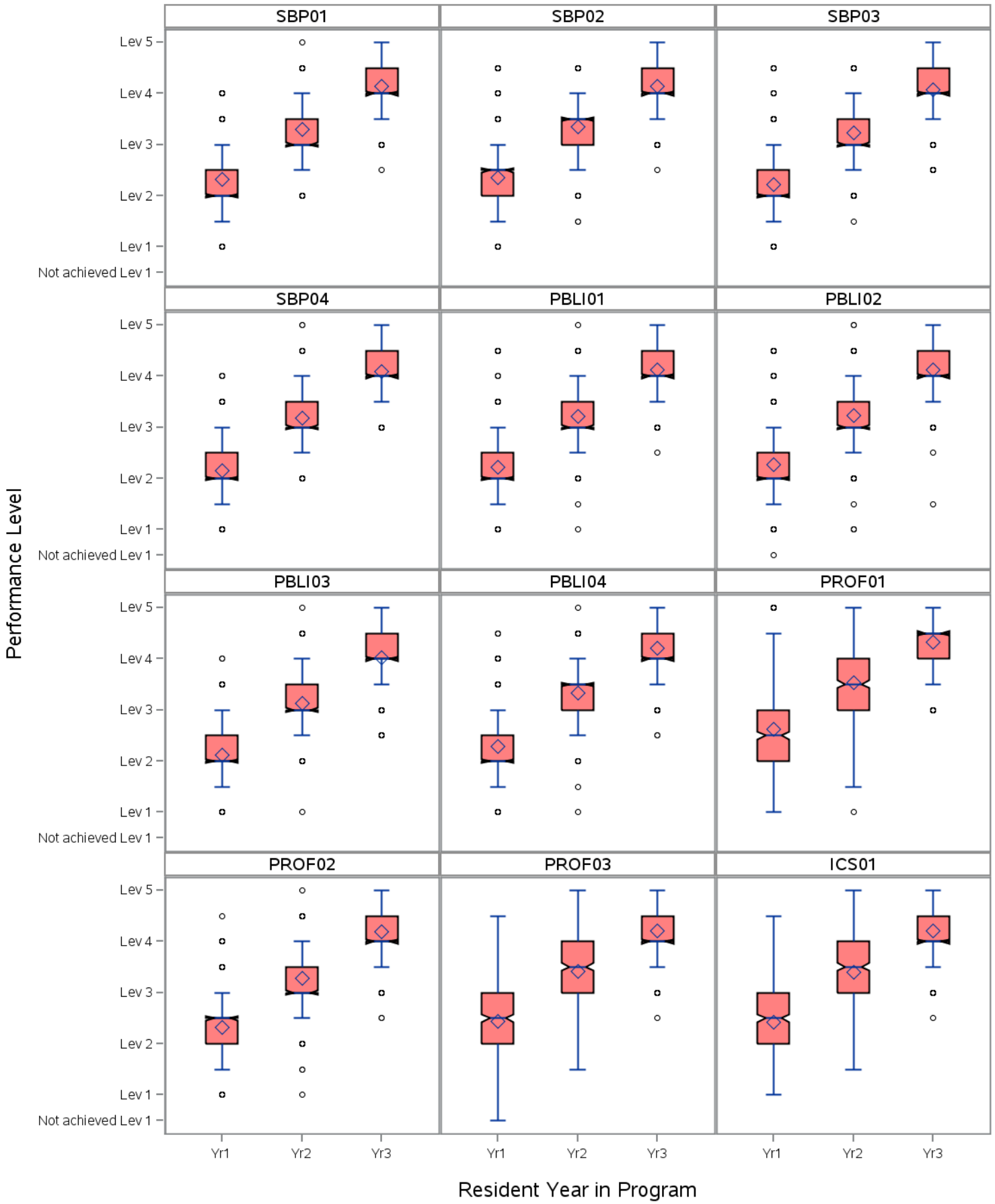


Table 6 - Dermatology (June 2016)

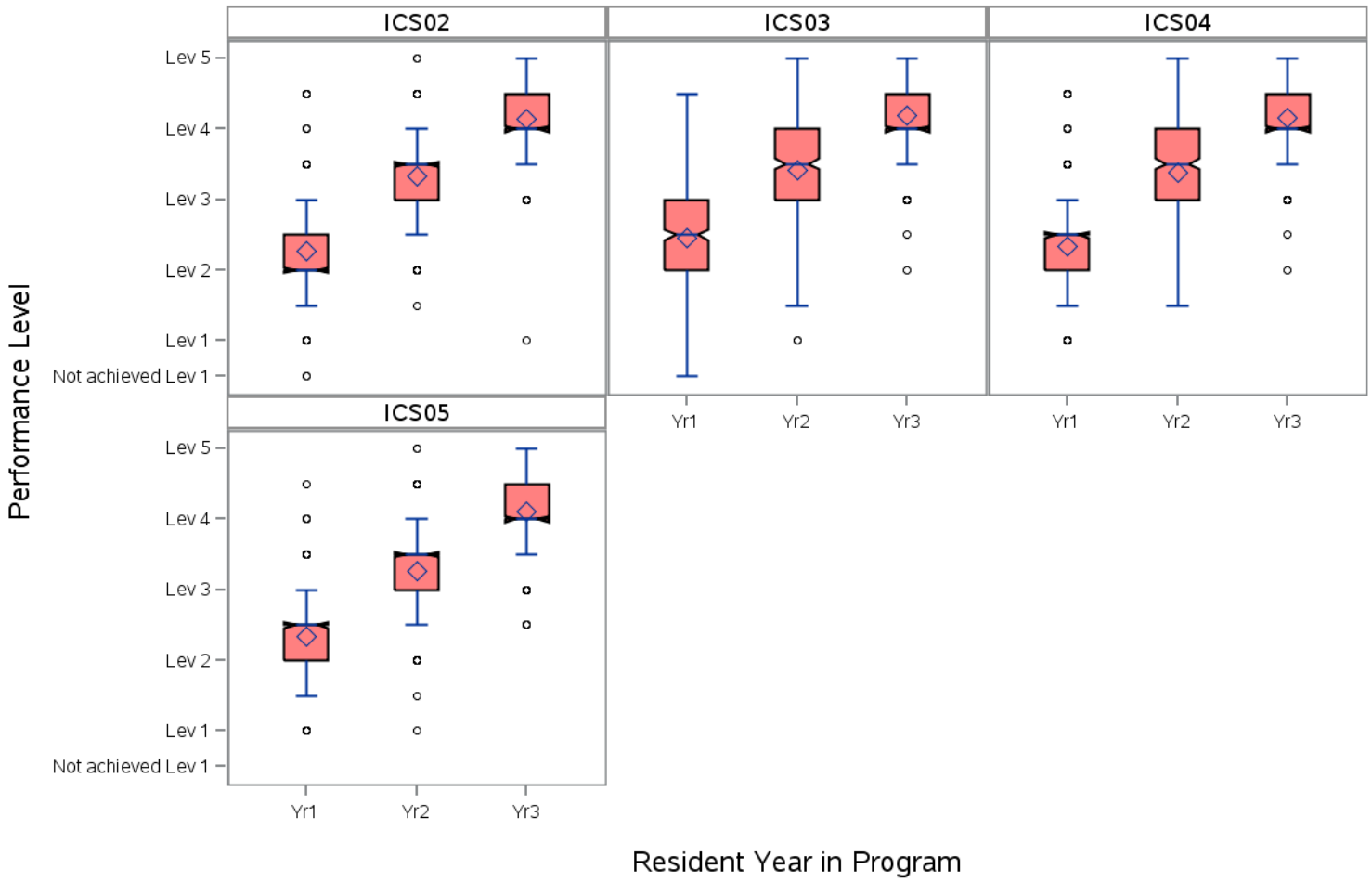


Table 7 - Emergency Medicine (June 2016)

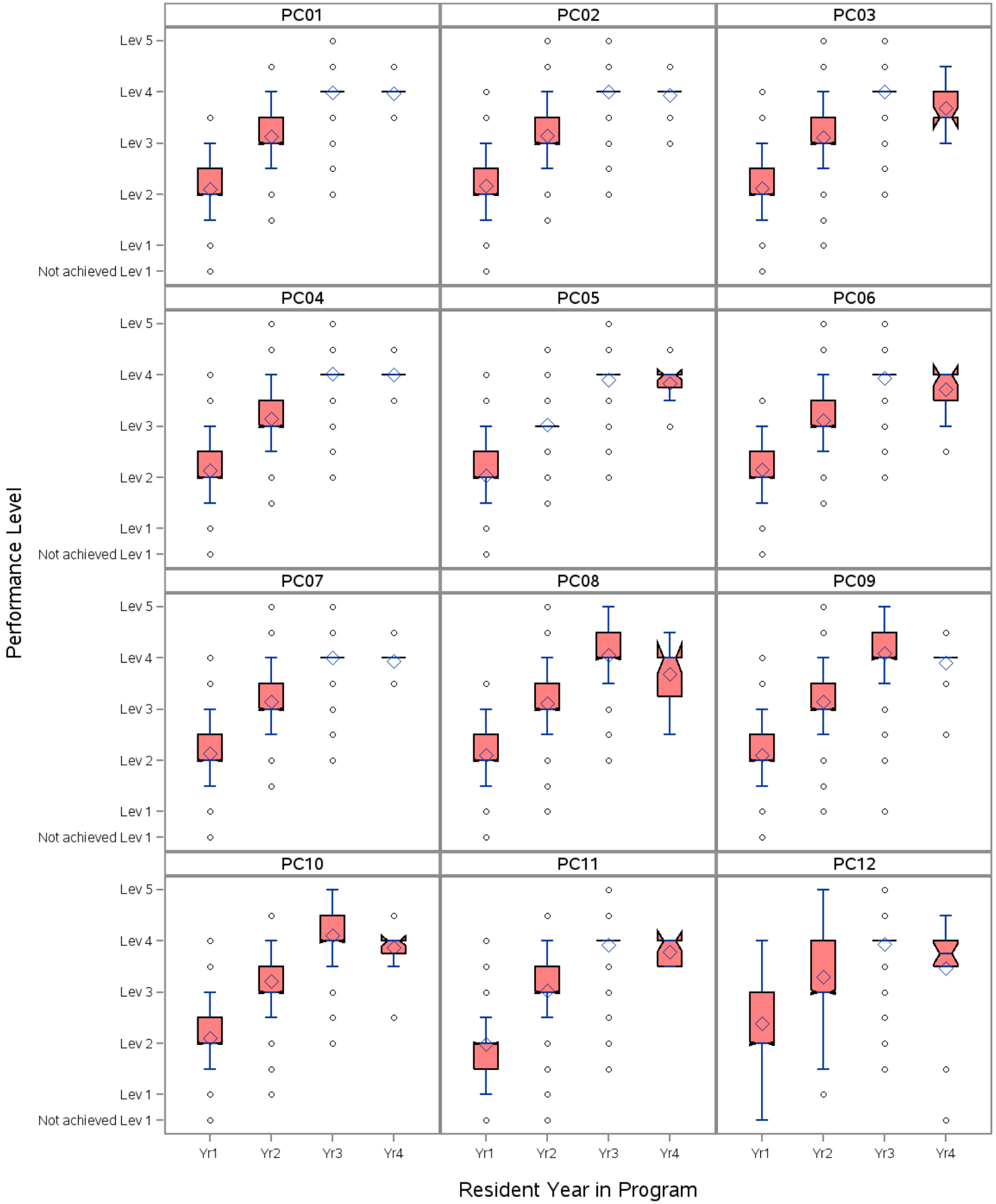


Table 7 - Emergency Medicine (June 2016)

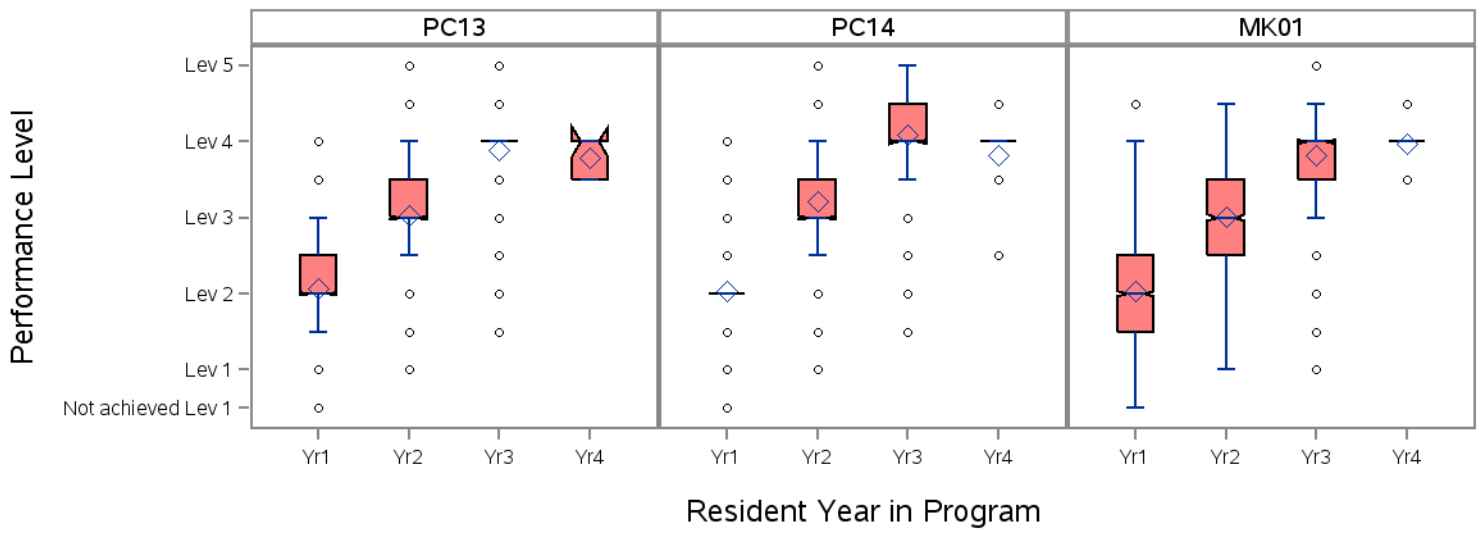


Table 7 - Emergency Medicine (June 2016)

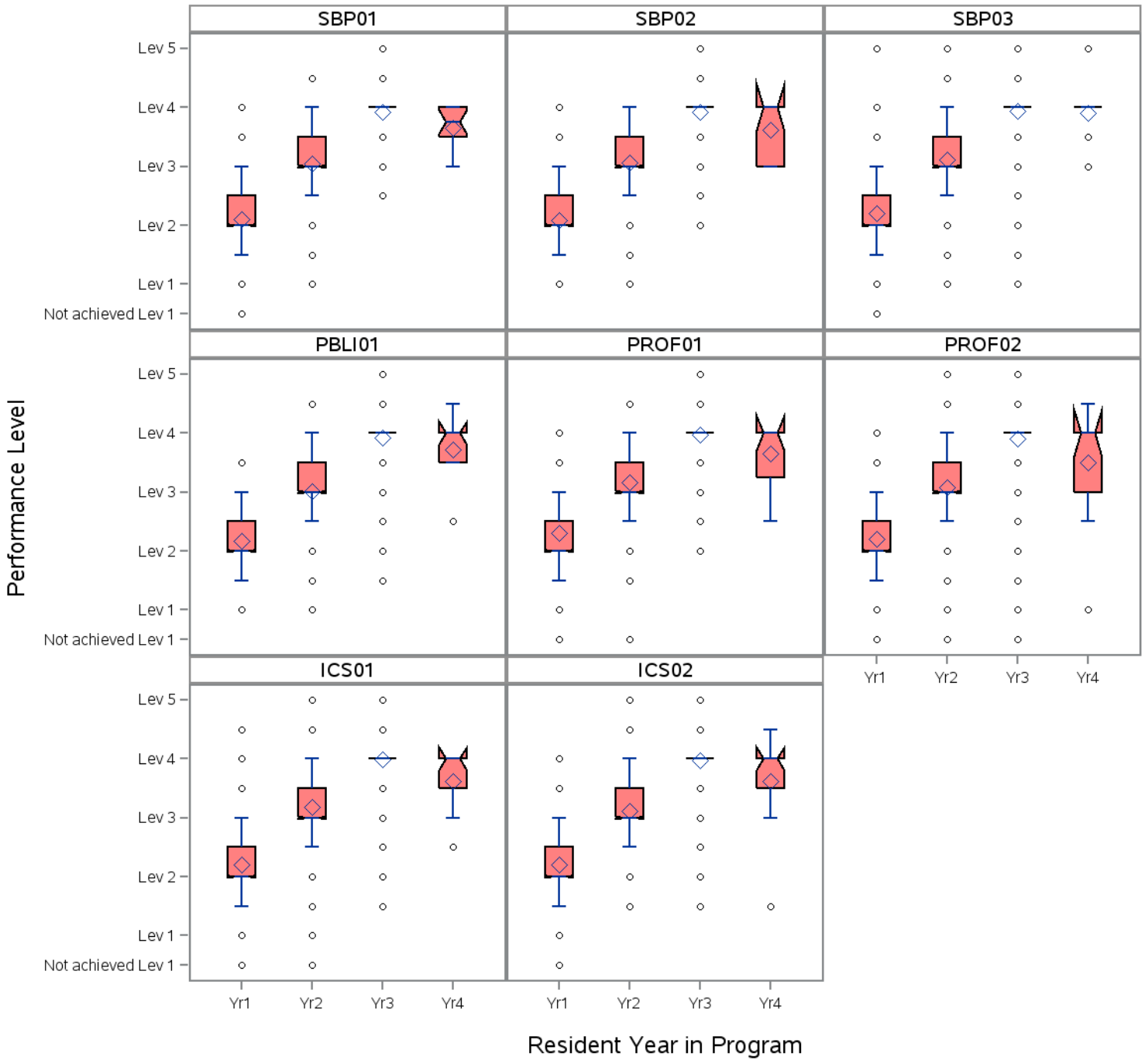


Table 8 - Family Medicine (June 2016)

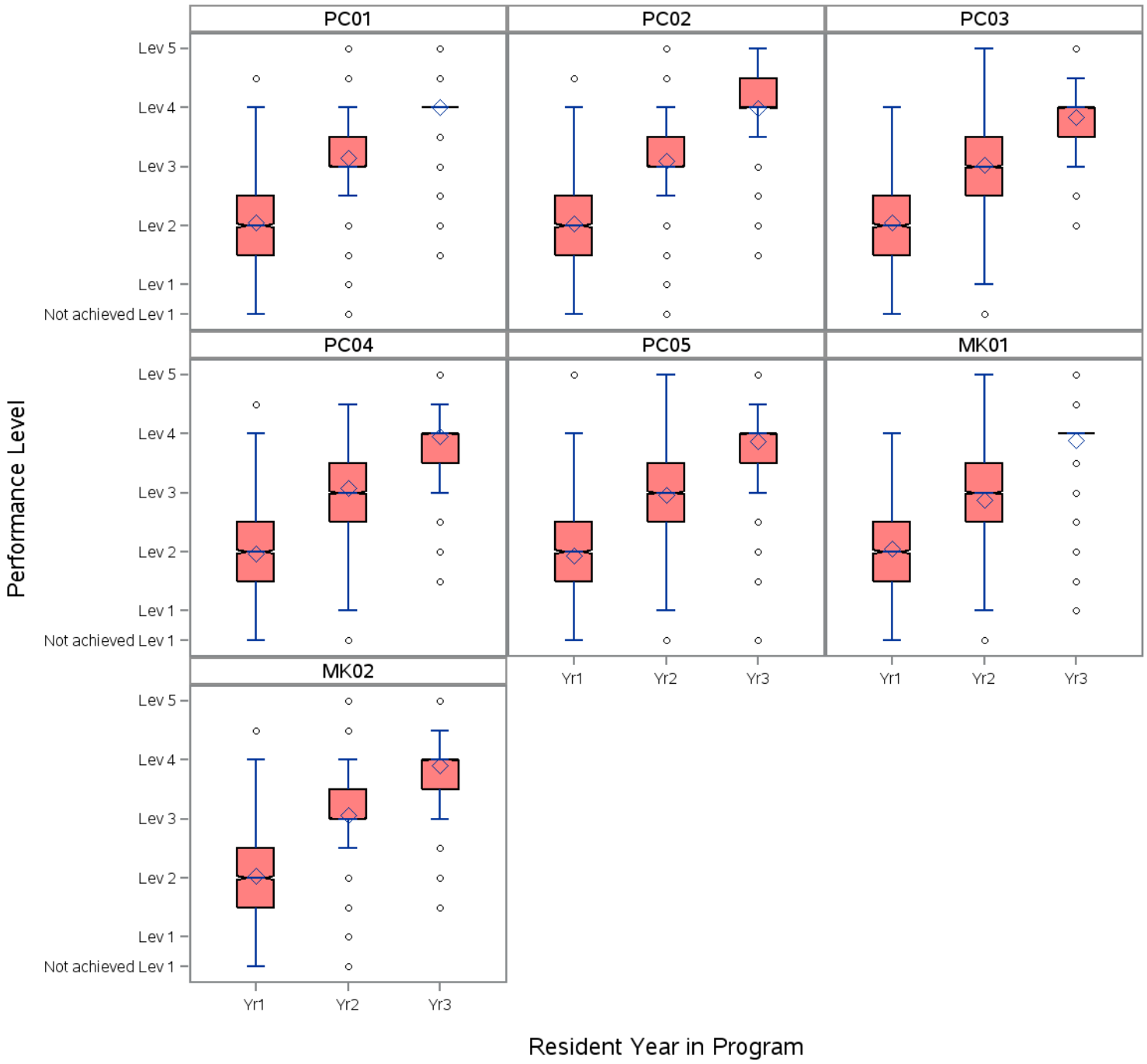


Table 8 - Family Medicine (June 2016)

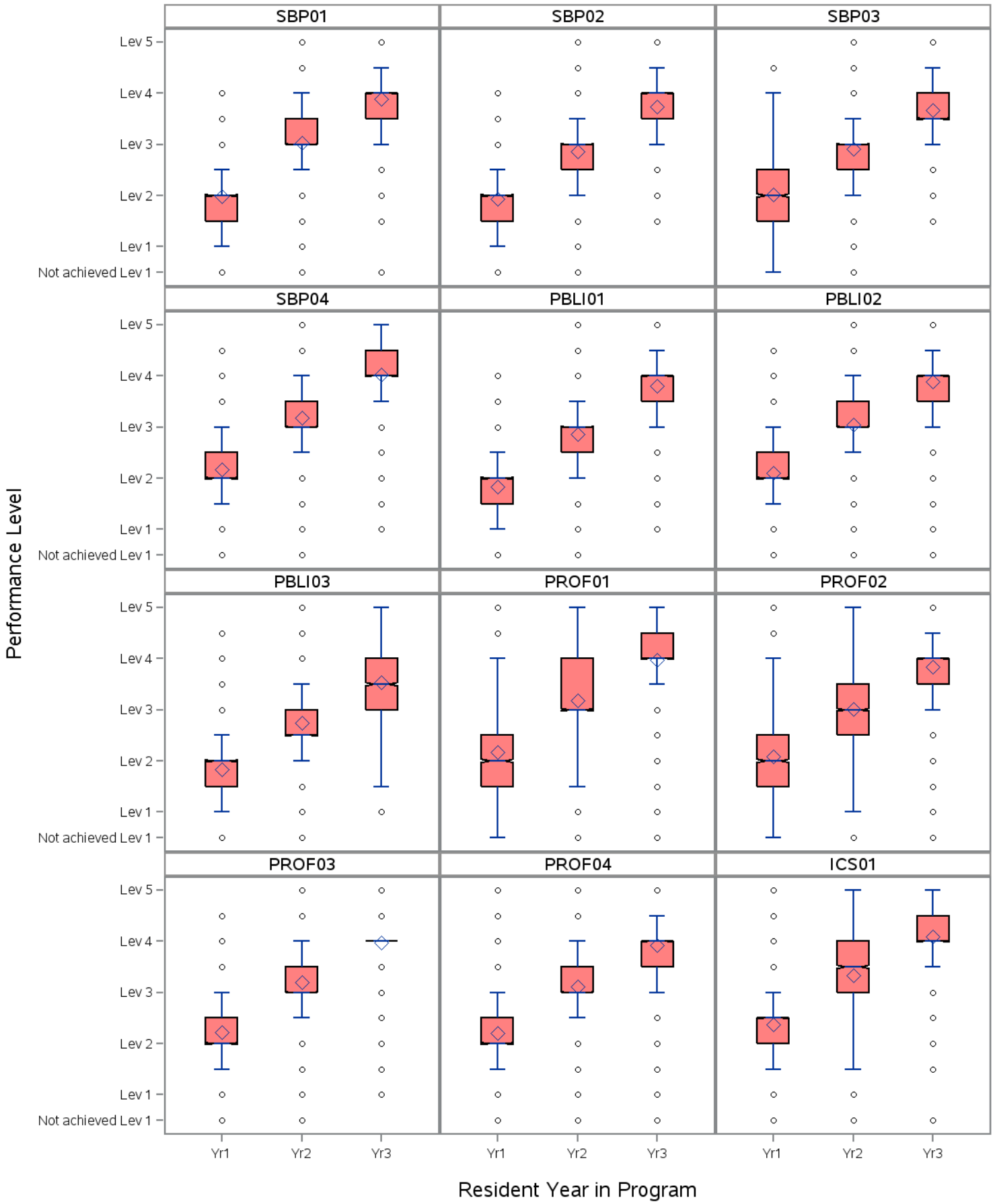


Table 8 - Family Medicine (June 2016)

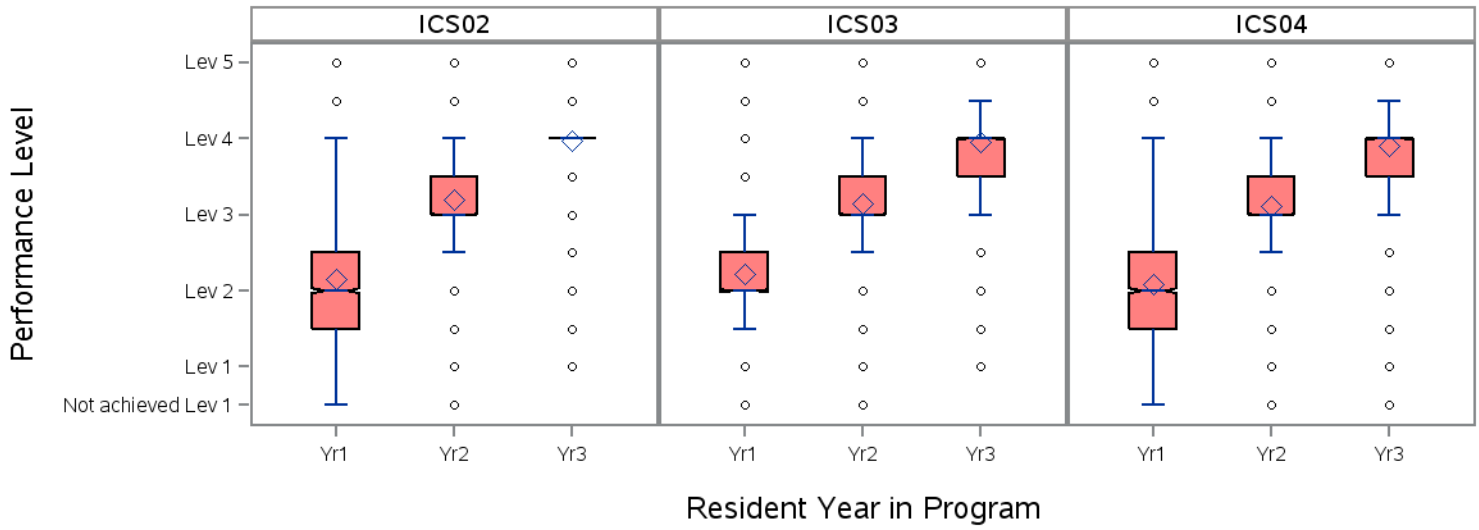


Table 9 - Internal Medicine (June 2016)

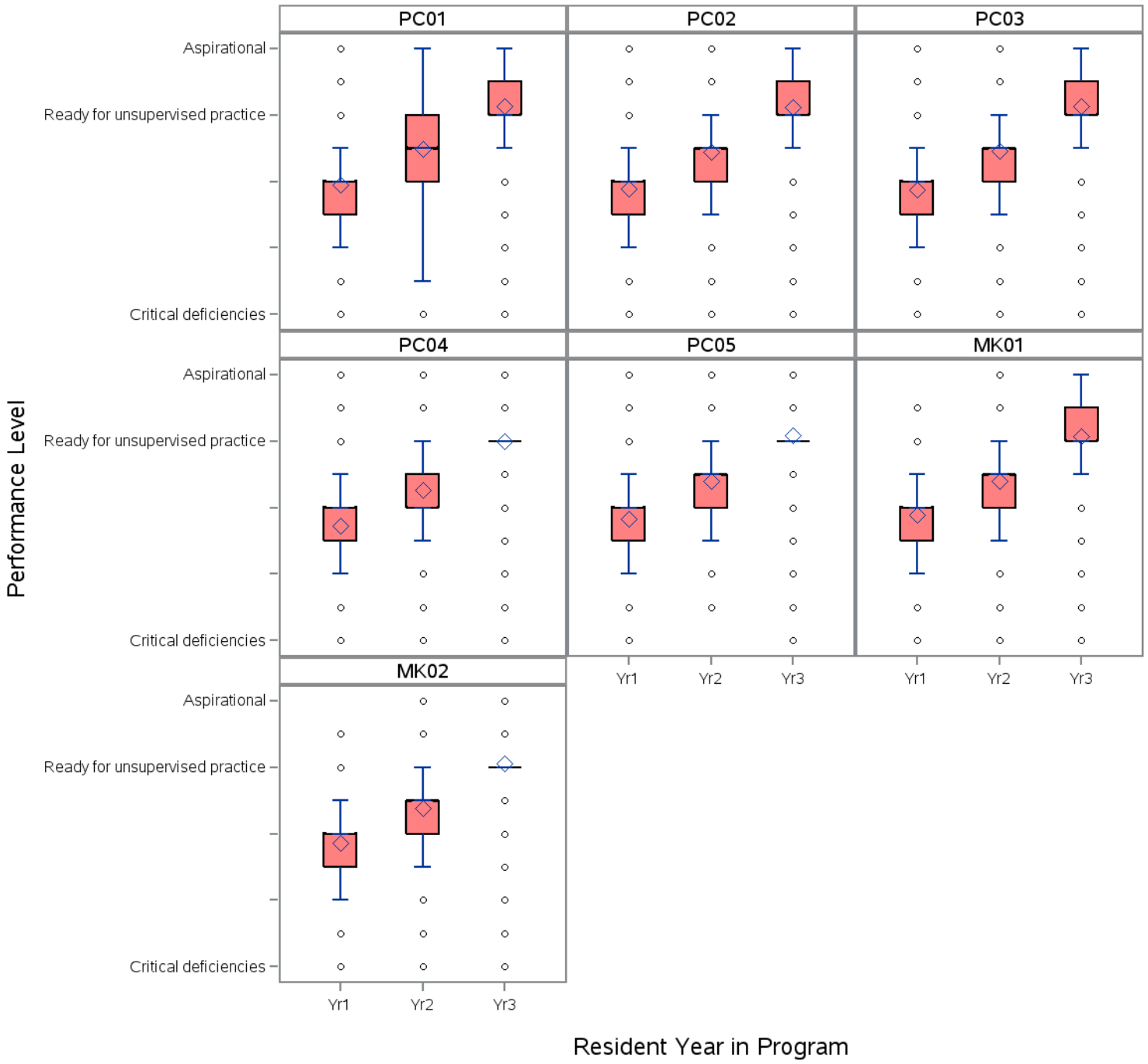


Table 9 - Internal Medicine (June 2016)

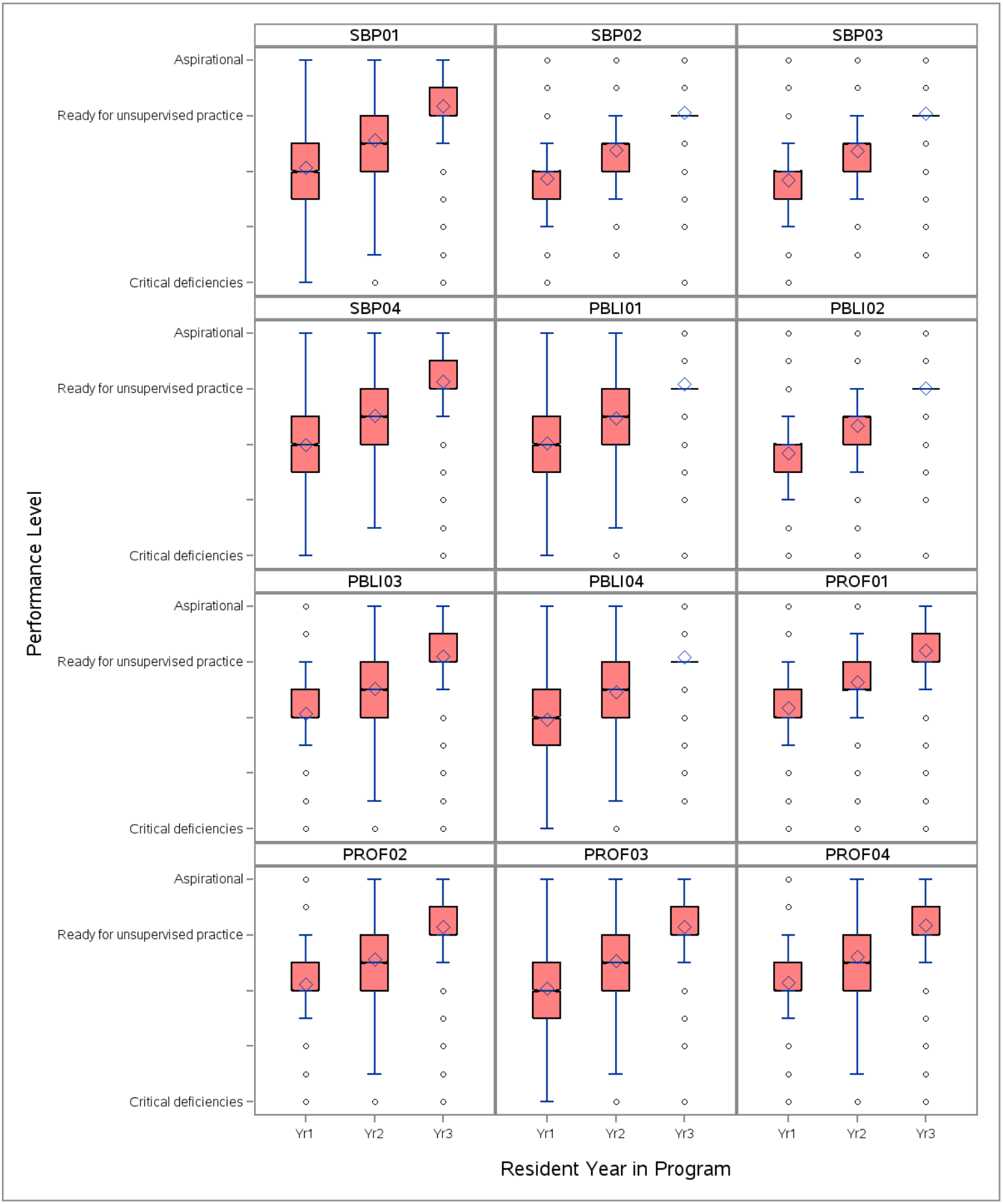


Table 9 - Internal Medicine (June 2016)

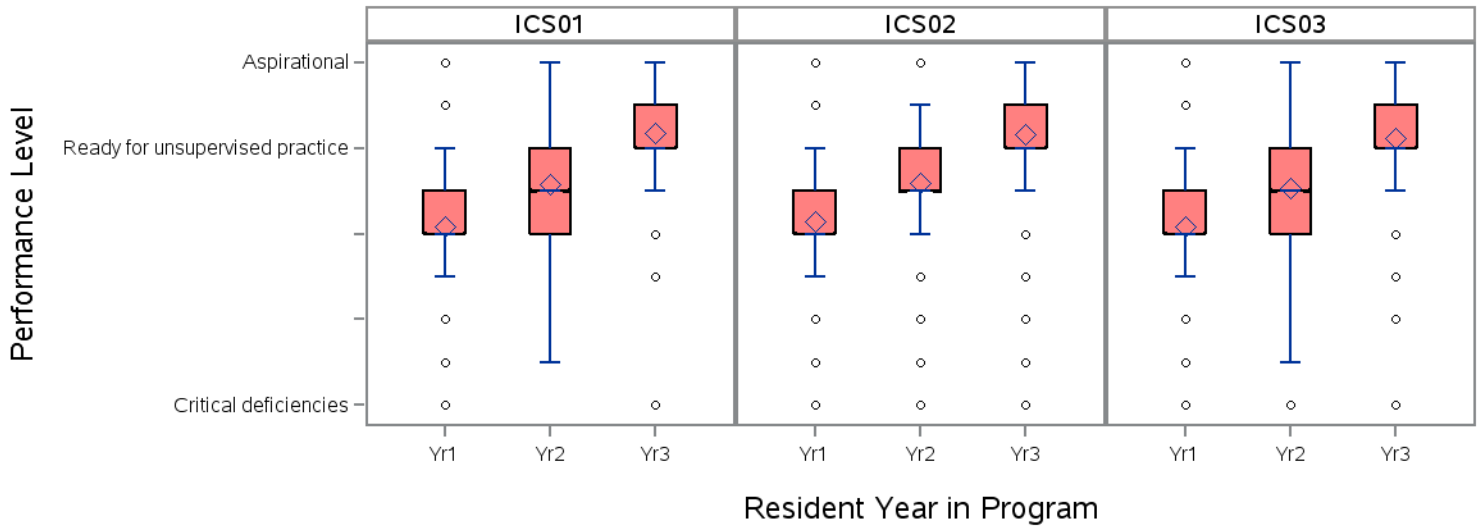


Table 10 - Internal Medicine/Pediatrics (June 2016)

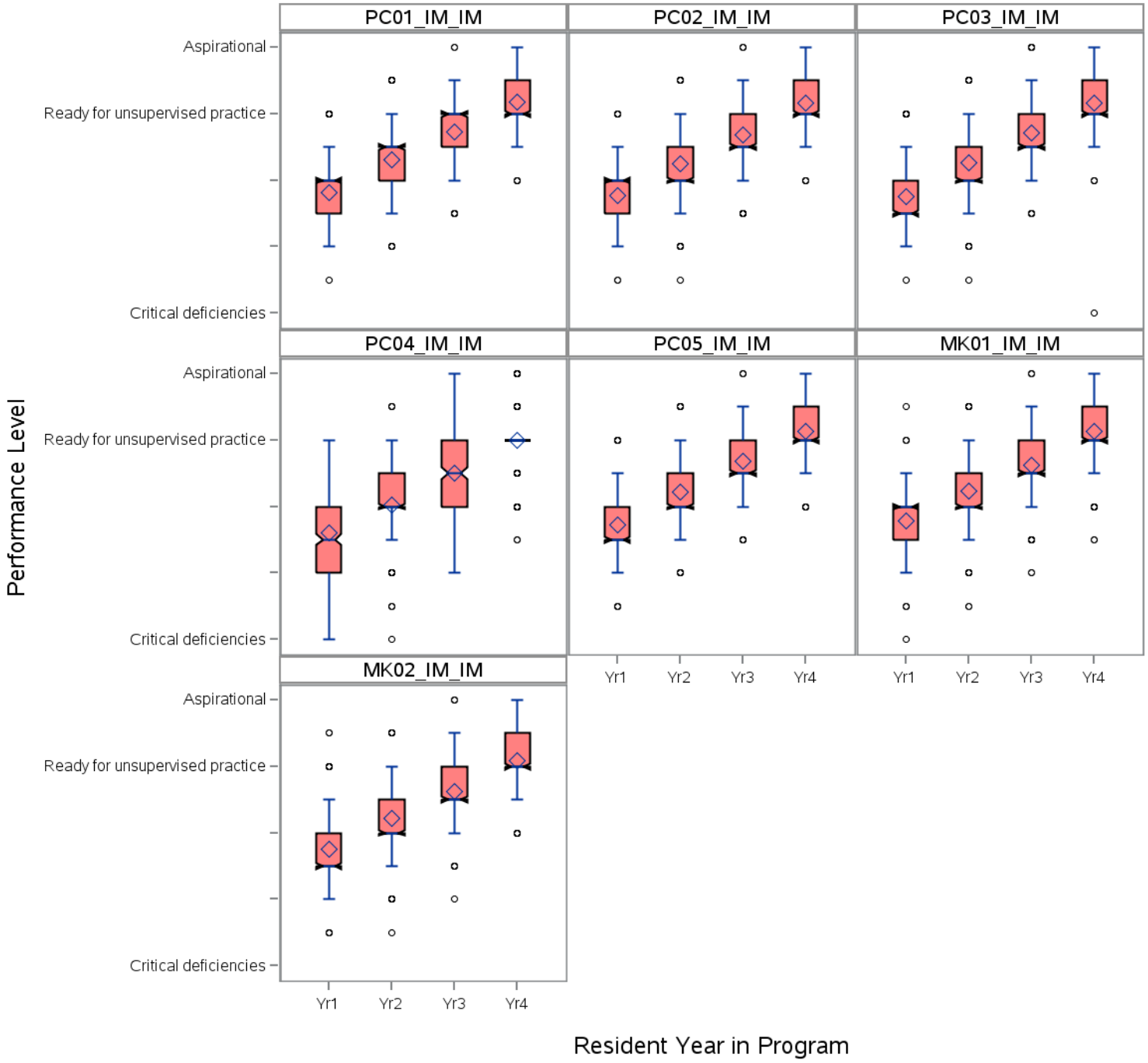


Table 10 - Internal Medicine/Pediatrics (June 2016)

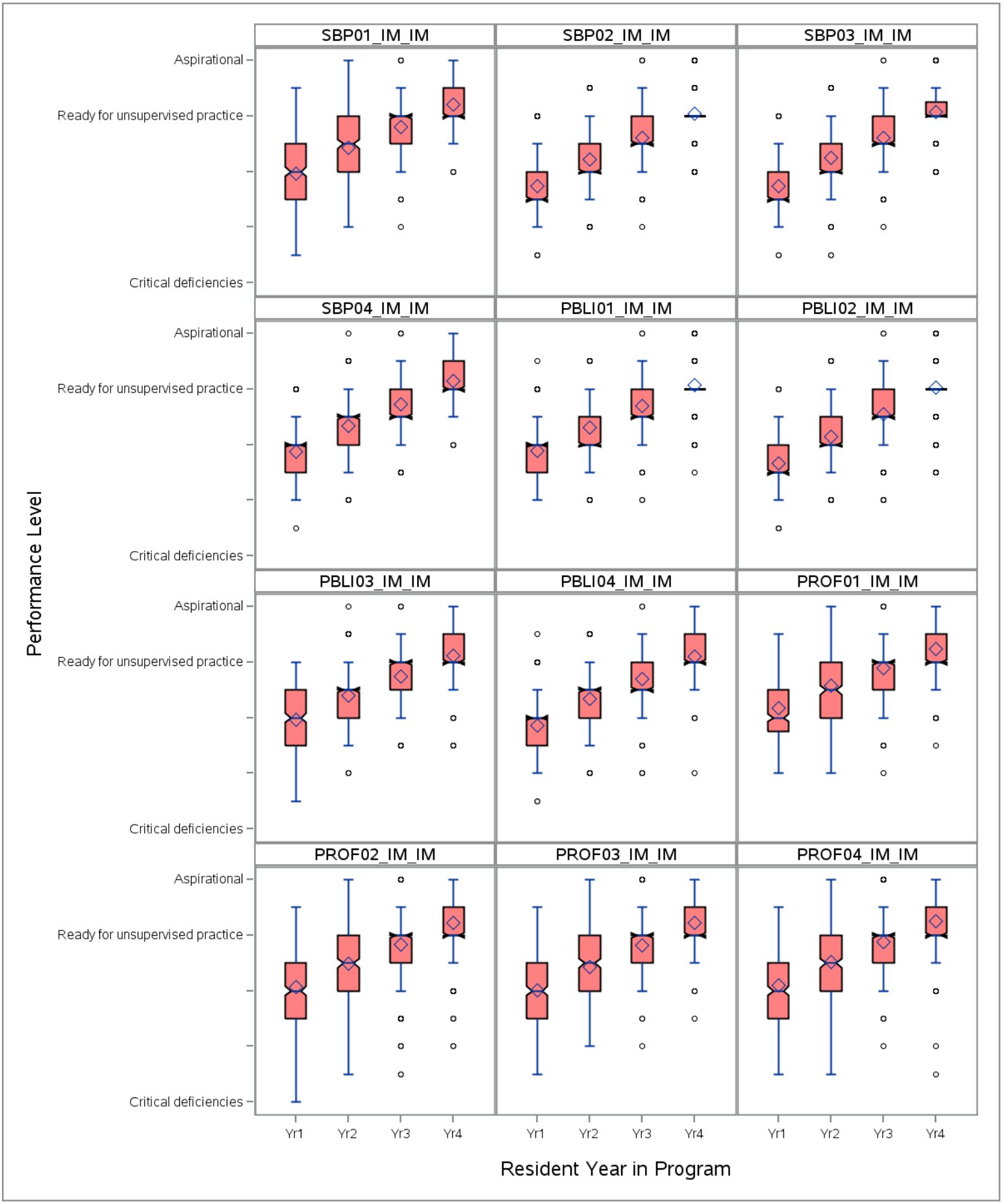


Table 10 - Internal Medicine/Pediatrics (June 2016)

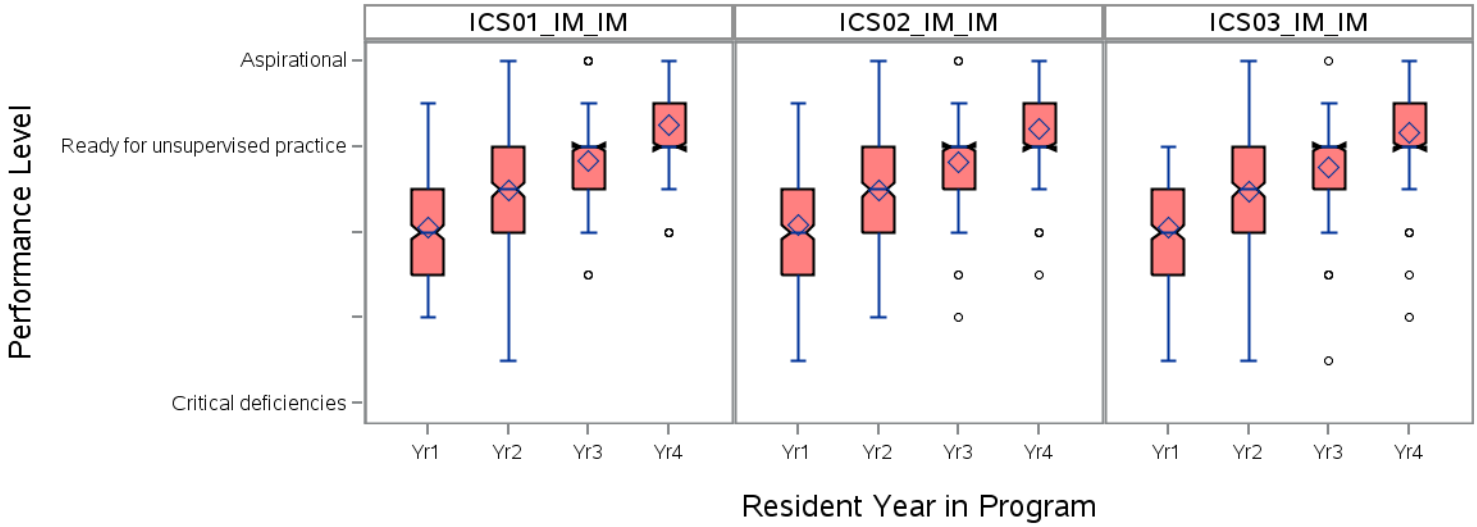


Table 10 - Internal Medicine/Pediatrics (June 2016)

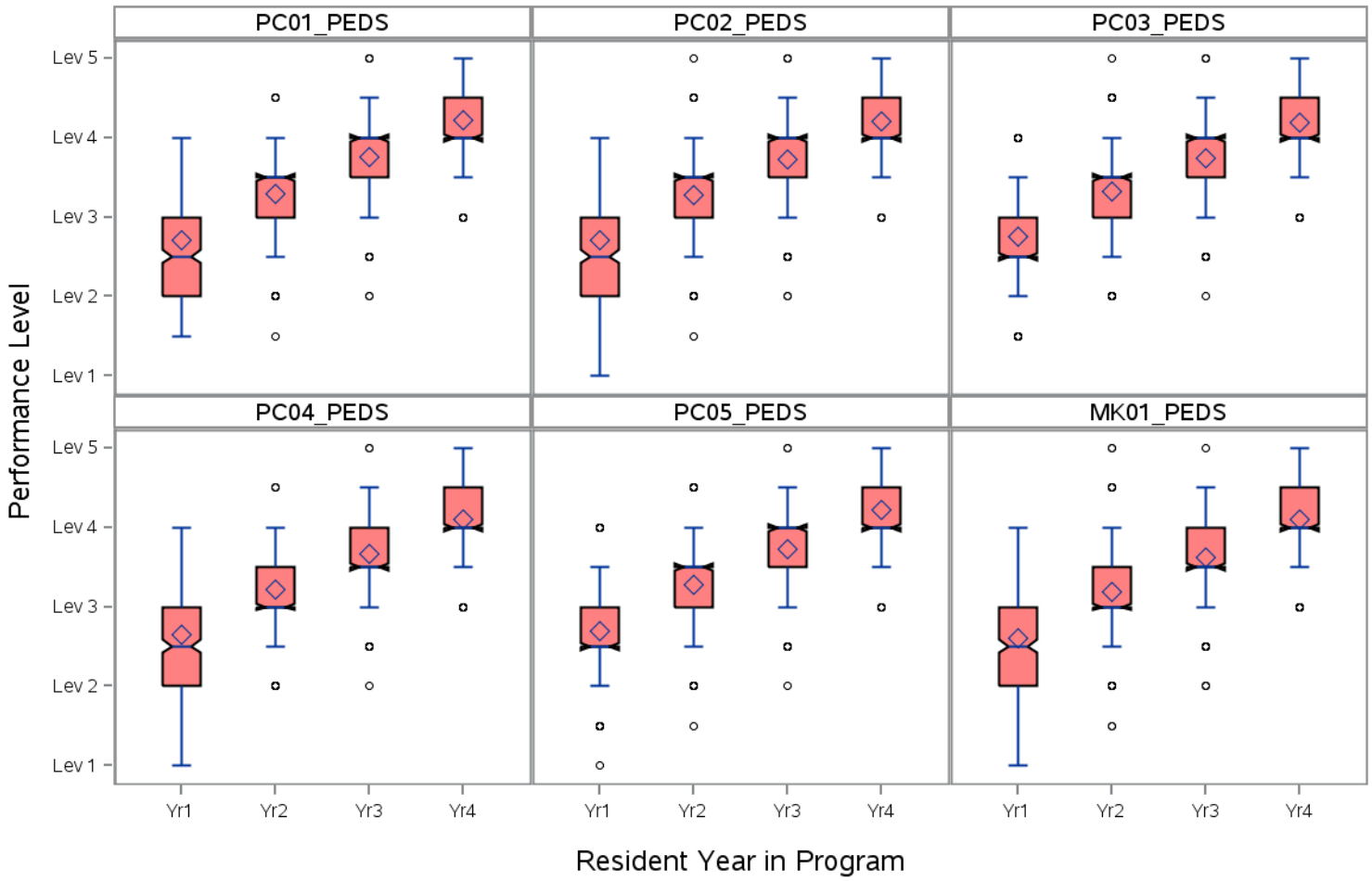


Table 10 - Internal Medicine/Pediatrics (June 2016)

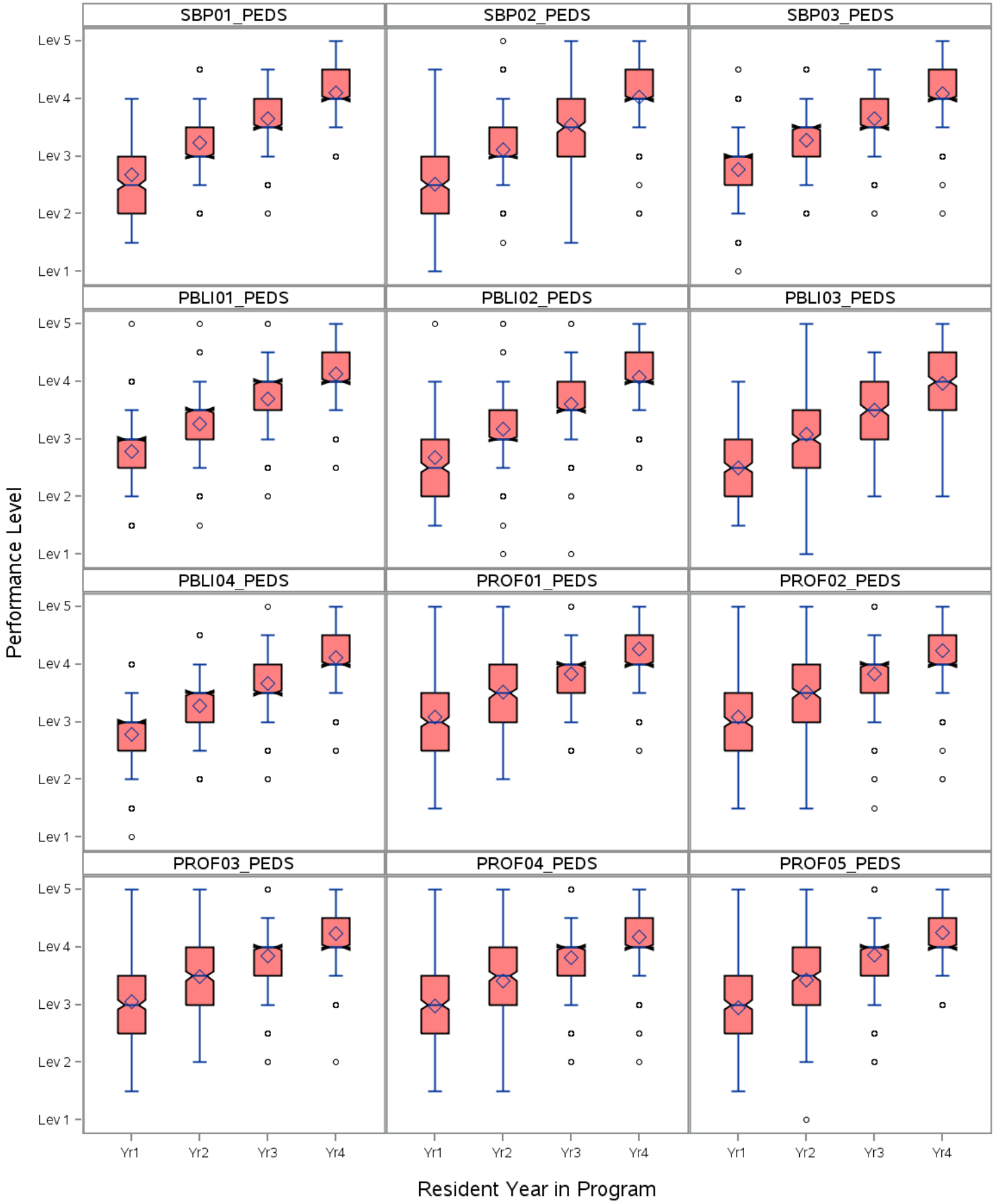


Table 10 - Internal Medicine/Pediatrics (June 2016)

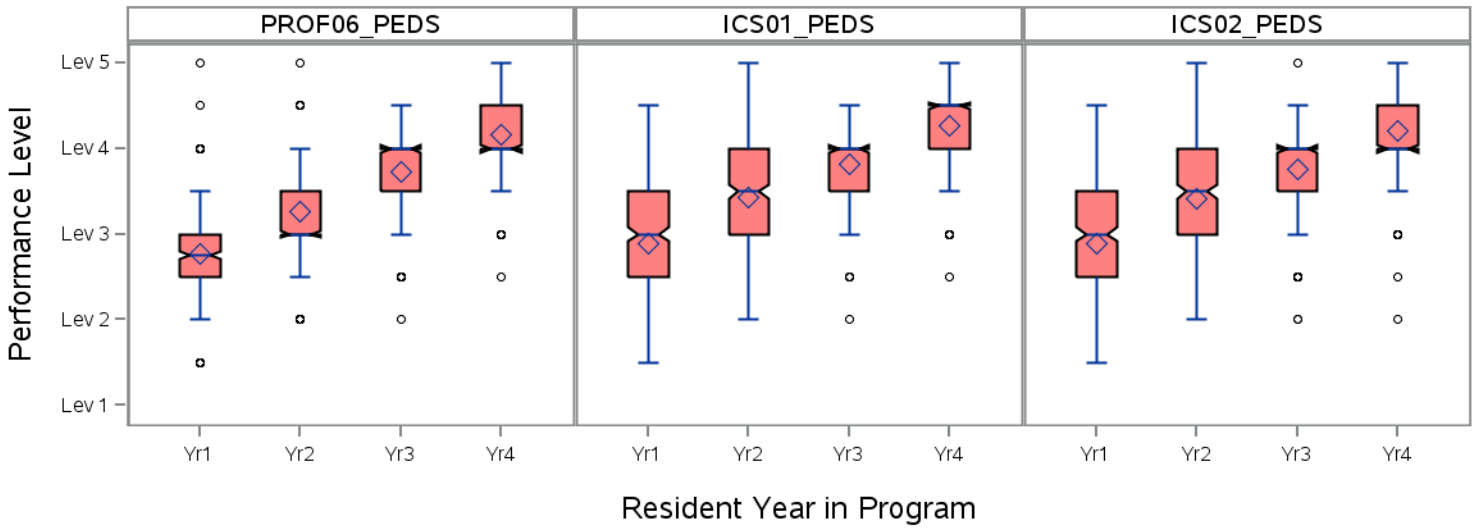


Table 11 - Medical Genetics and Genomics (June 2016)

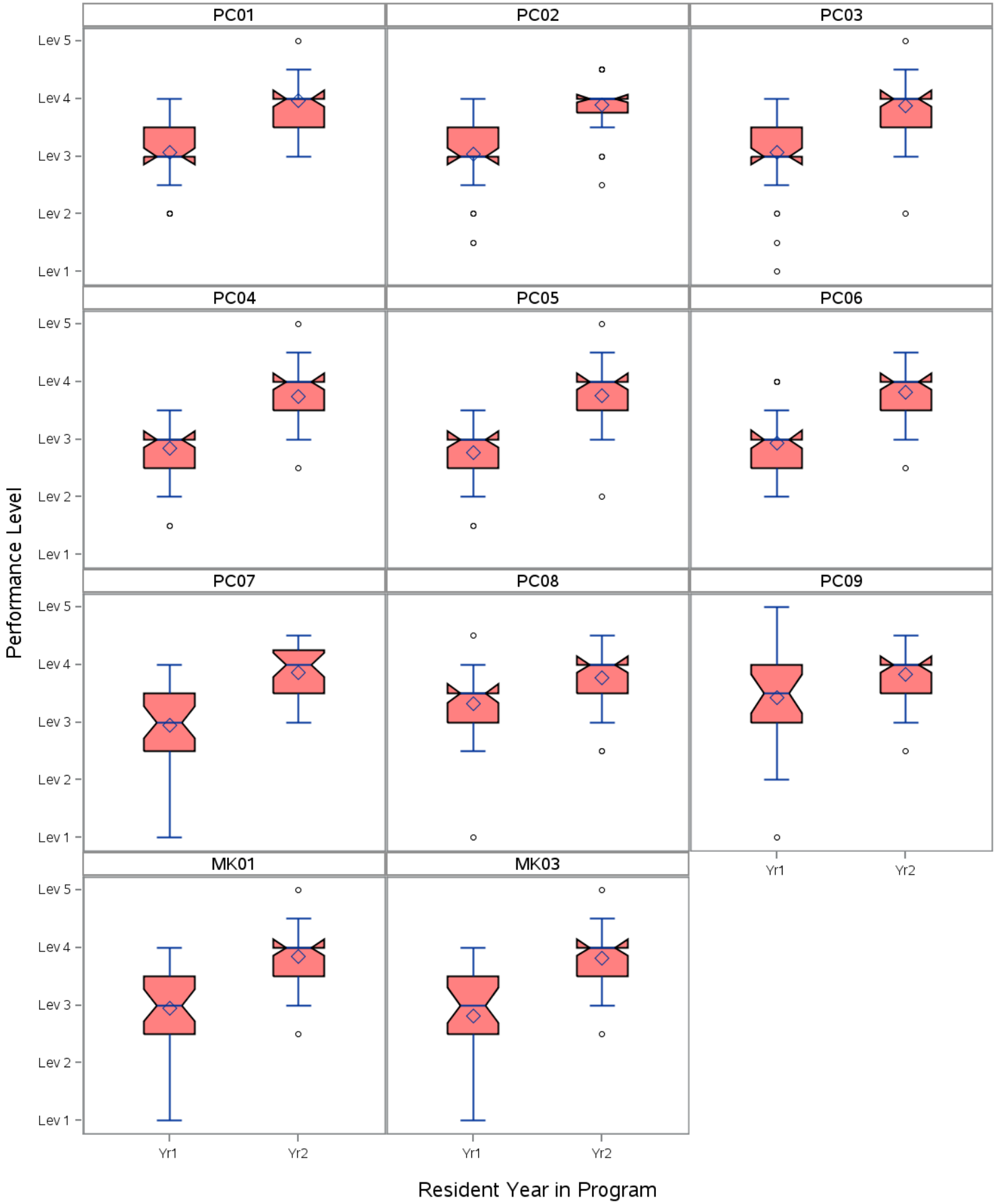


Table 11 - Medical Genetics and Genomics (June 2016)

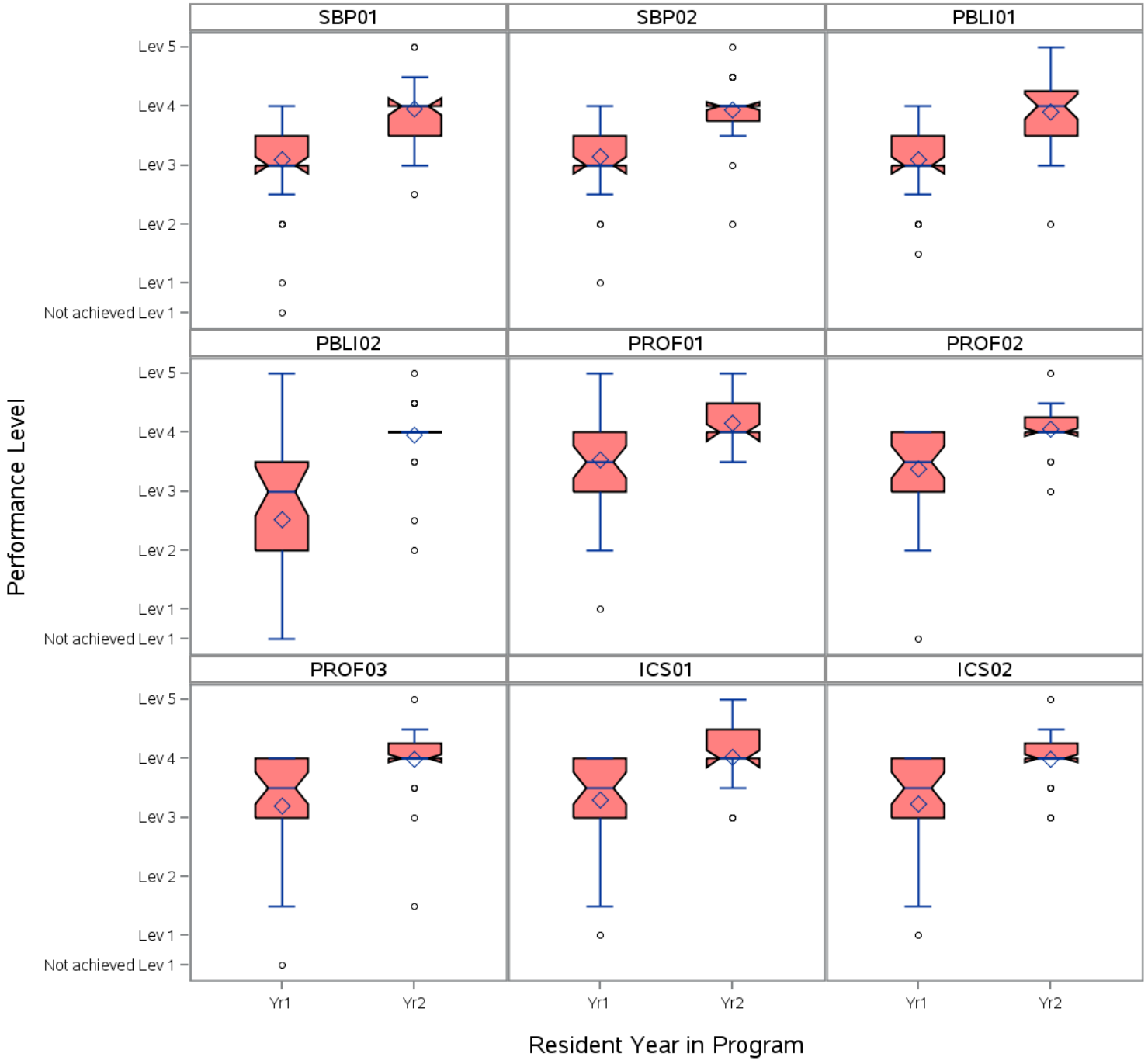


Table 12 - Neurological Surgery (June 2016)

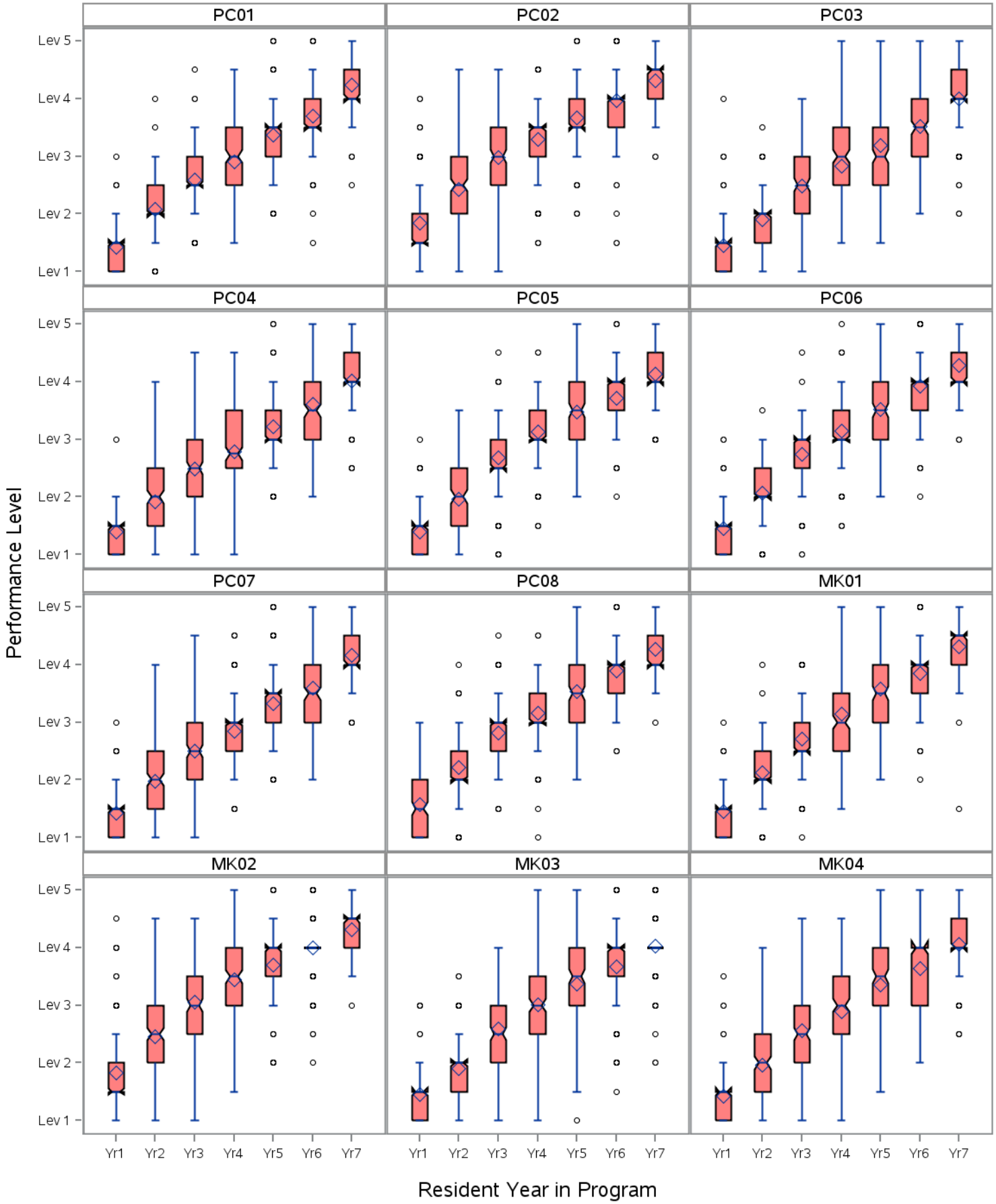


Table 12 - Neurological Surgery (June 2016)

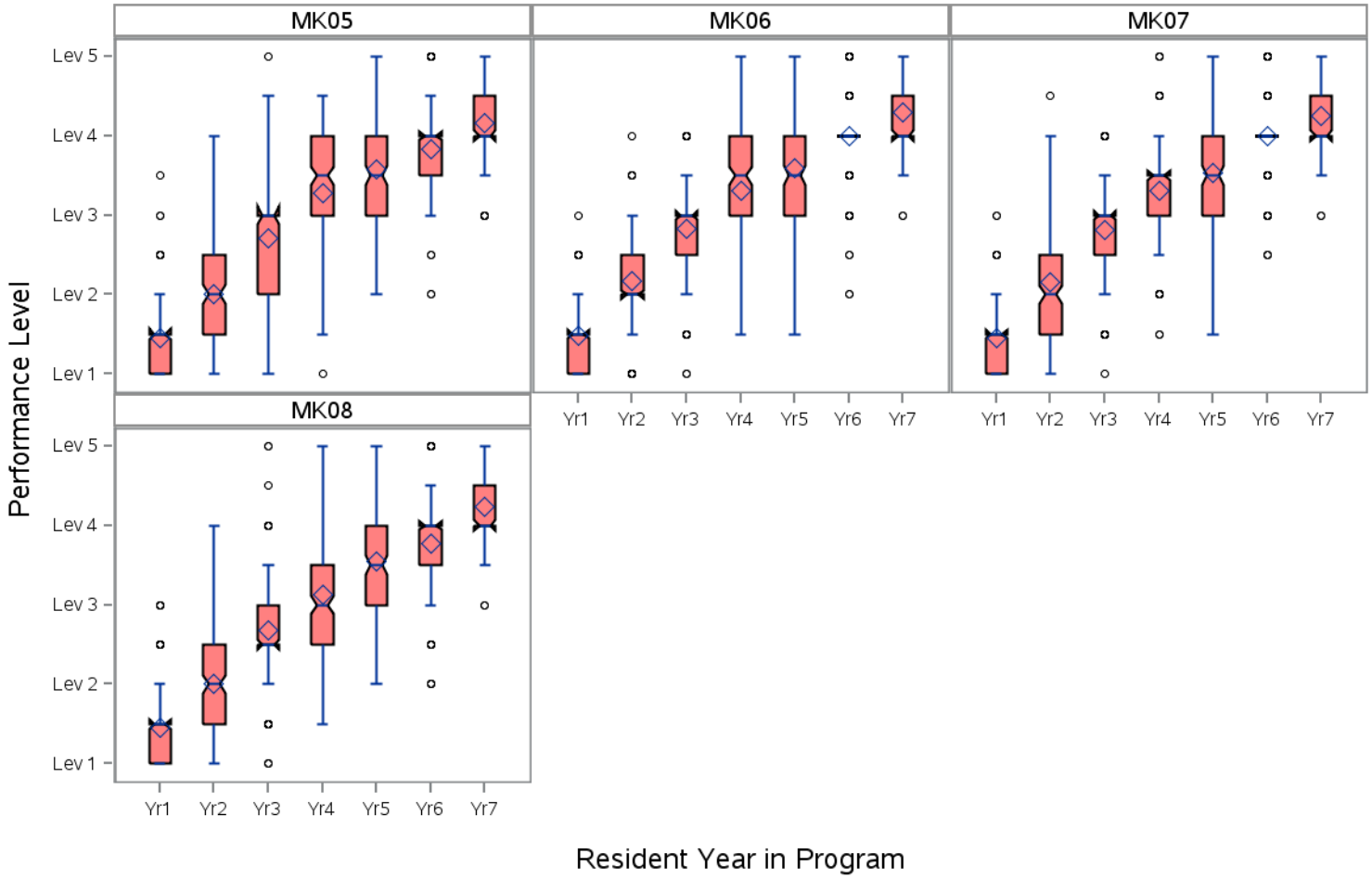


Table 12 - Neurological Surgery (June 2016)

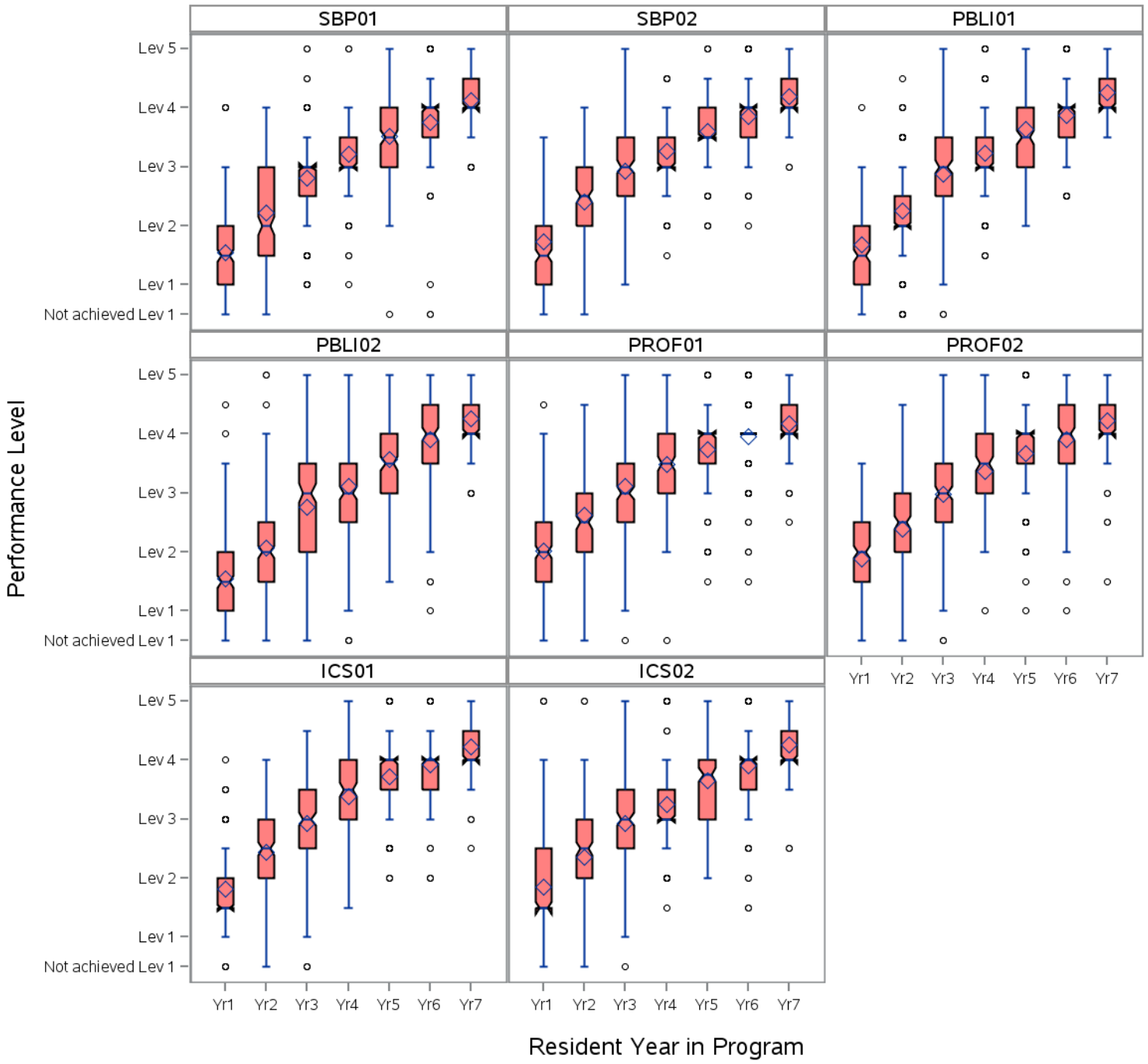


Table 13 - Neurology (June 2016)

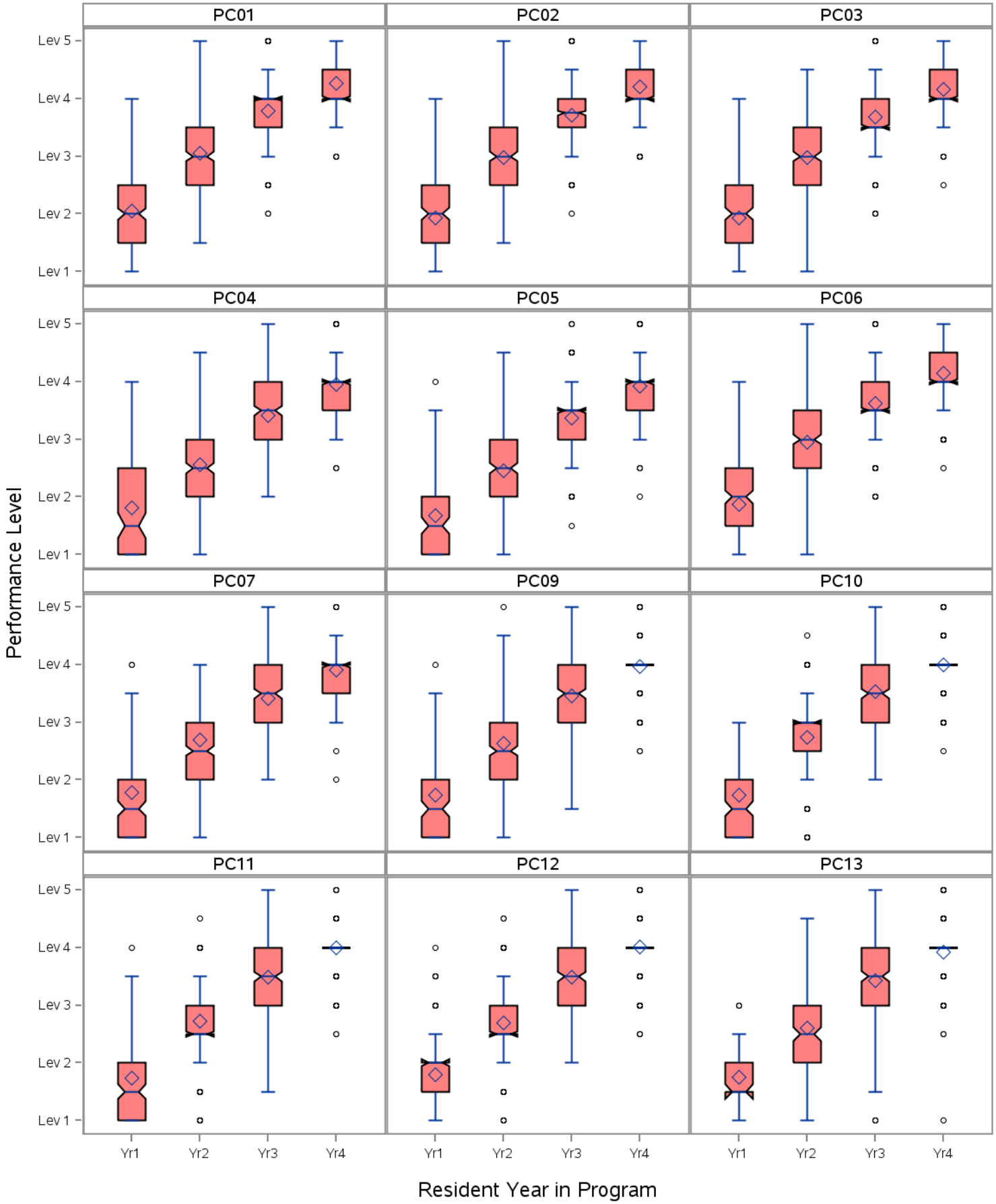


Table 13 - Neurology (June 2016)

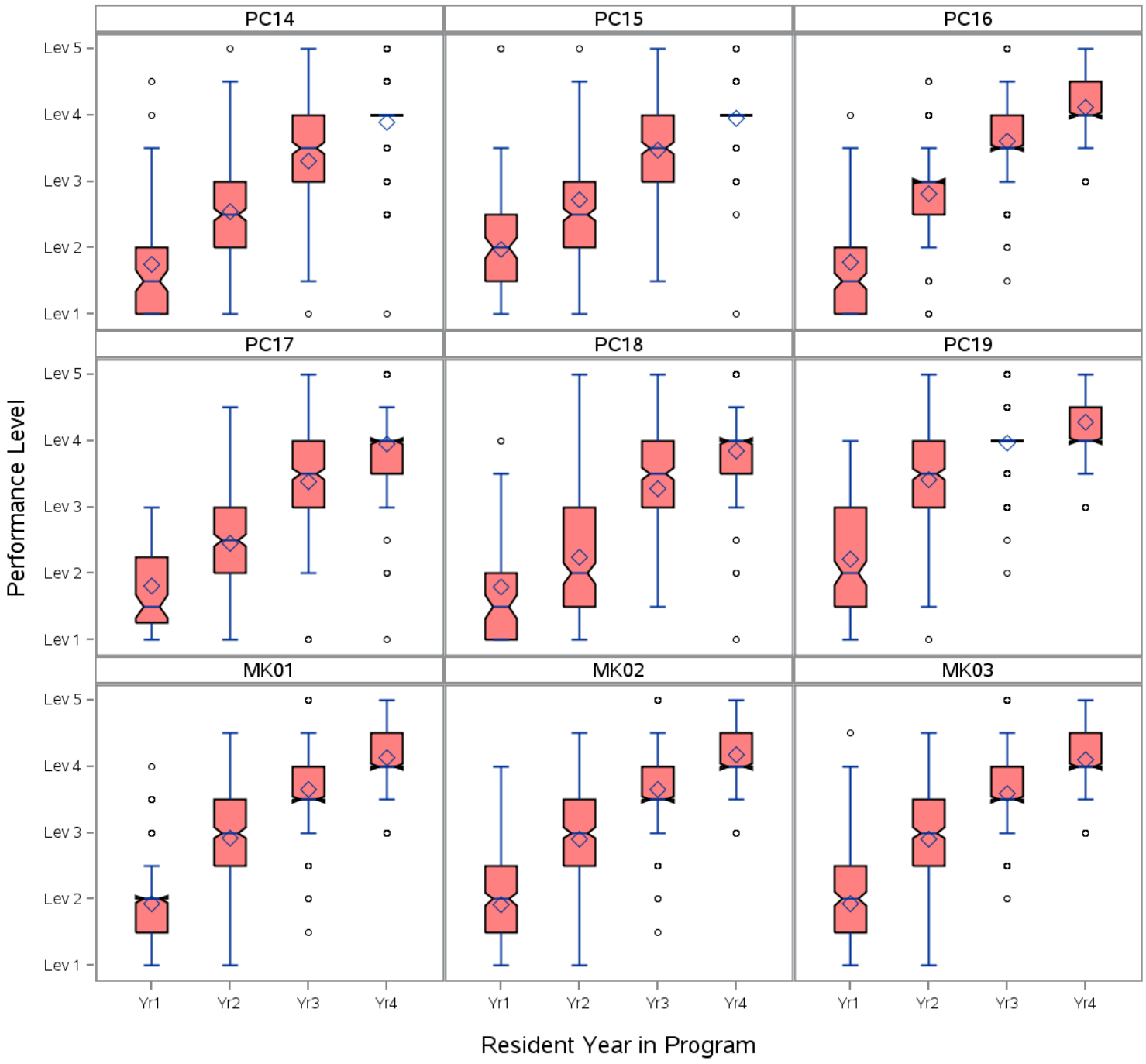


Table 13 - Neurology (June 2016)

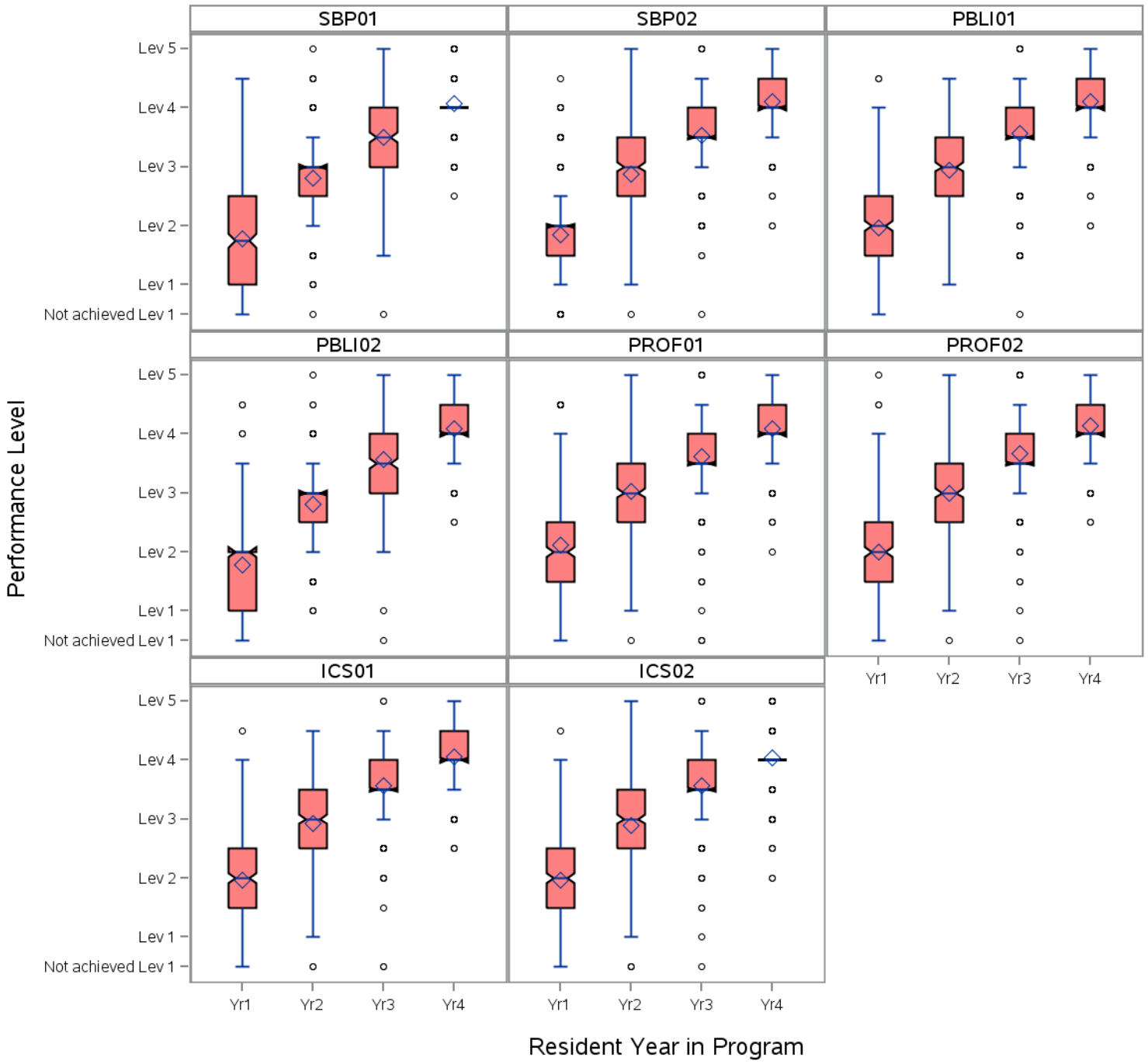


Table 14 - Nuclear Medicine (June 2016)

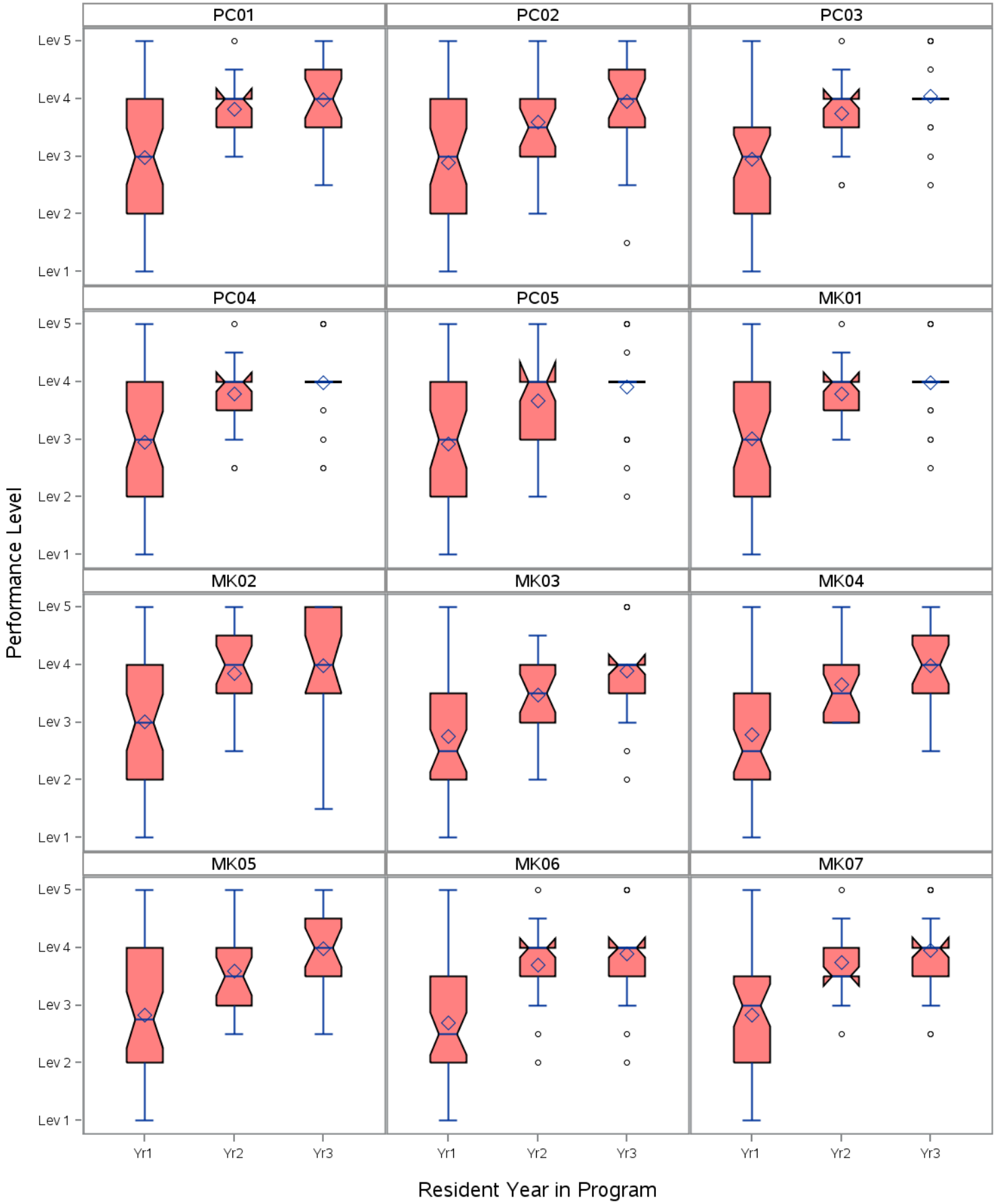


Table 14 - Nuclear Medicine (June 2016)

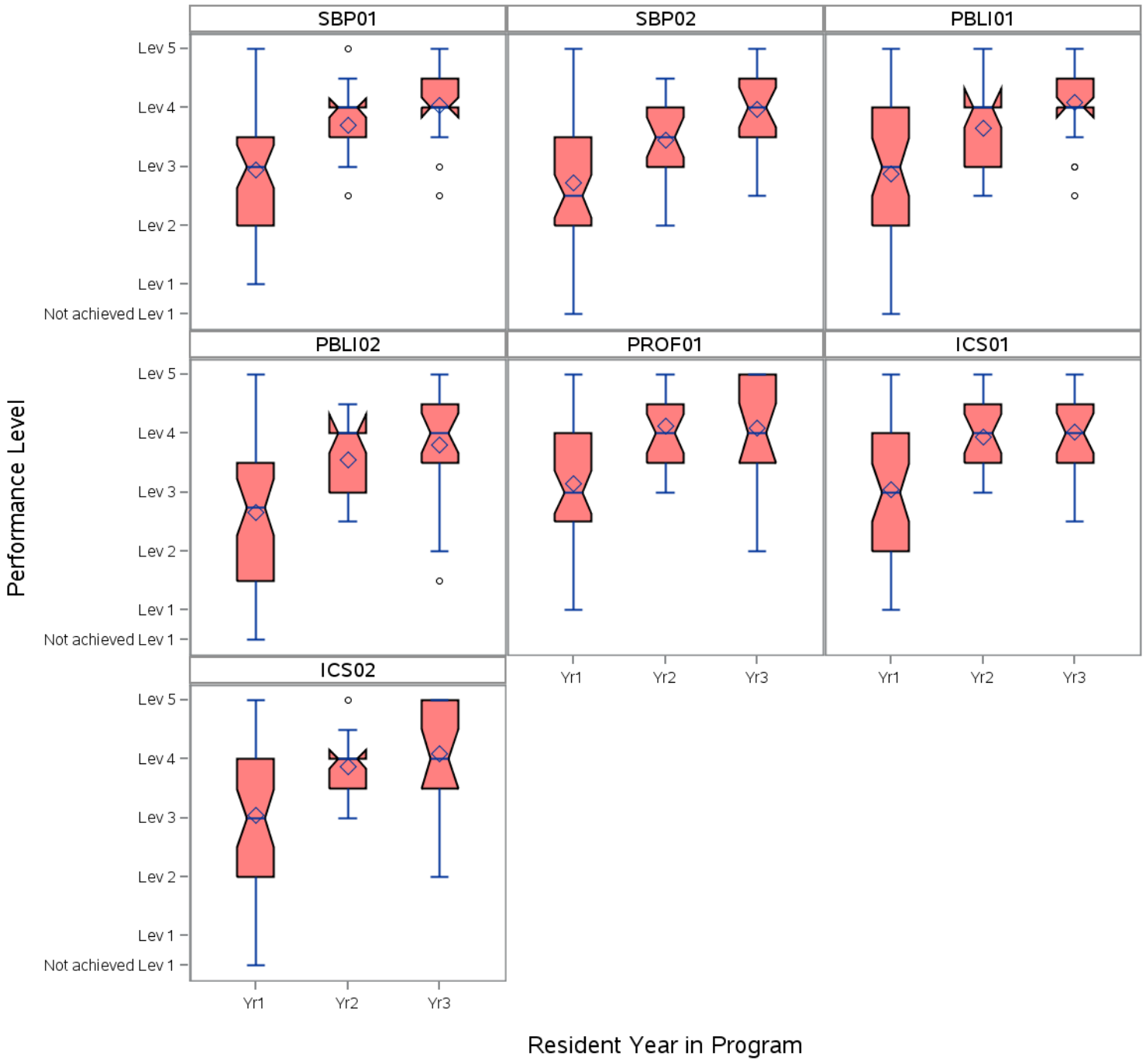


Table 15 - Obstetrics and Gynecology (June 2016)

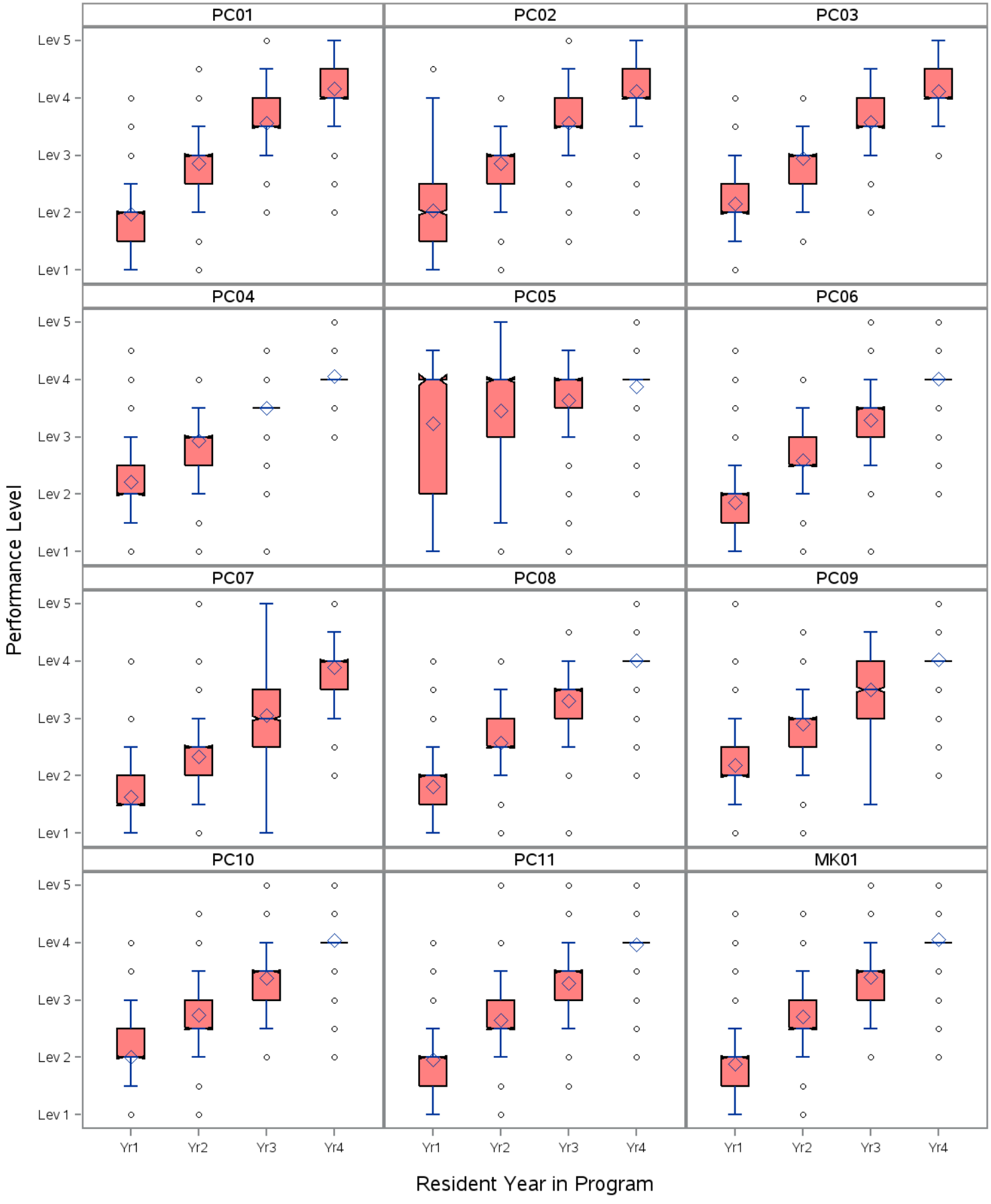


Table 15 - Obstetrics and Gynecology (June 2016)

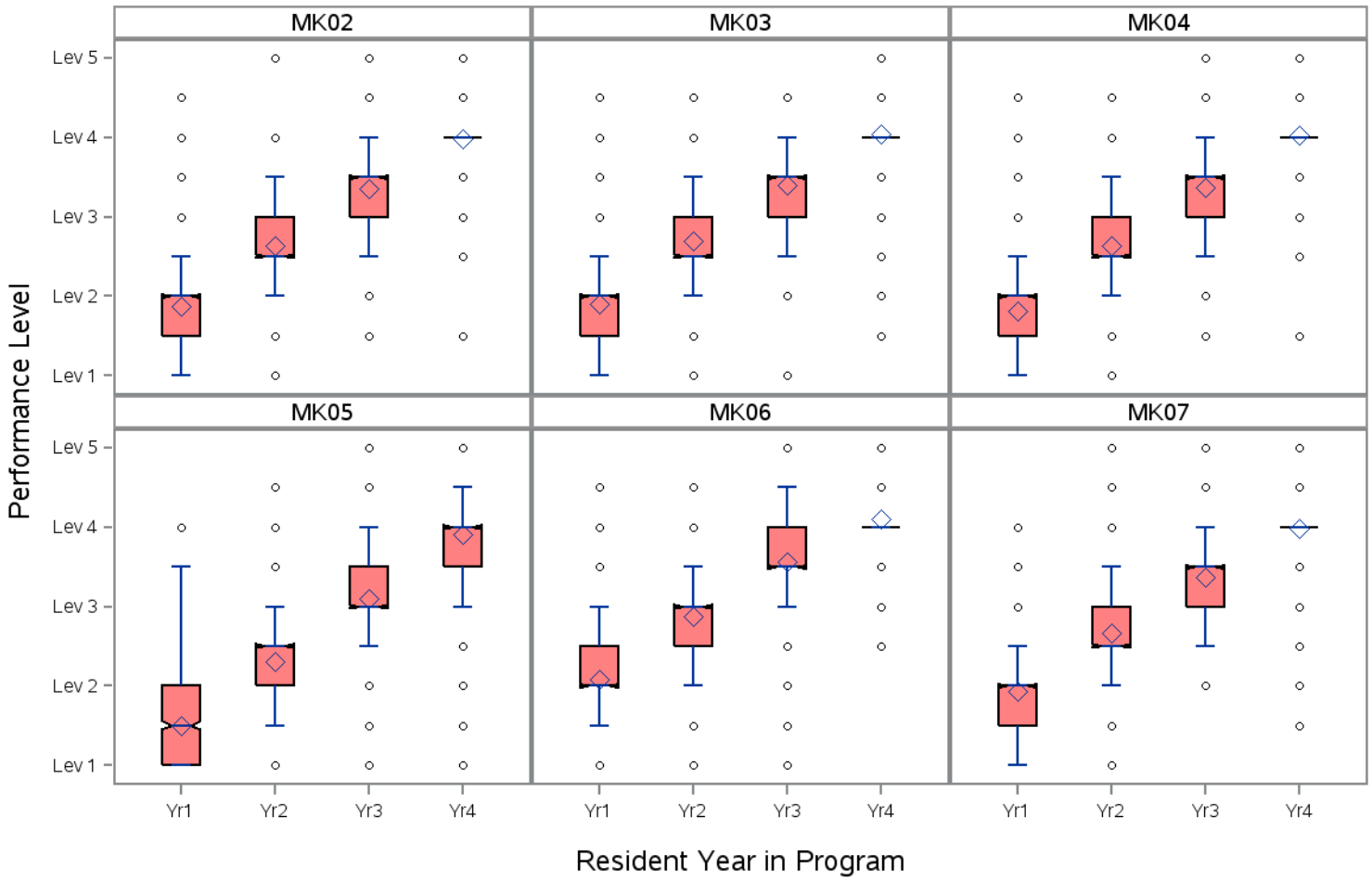


Table 15 - Obstetrics and Gynecology (June 2016)

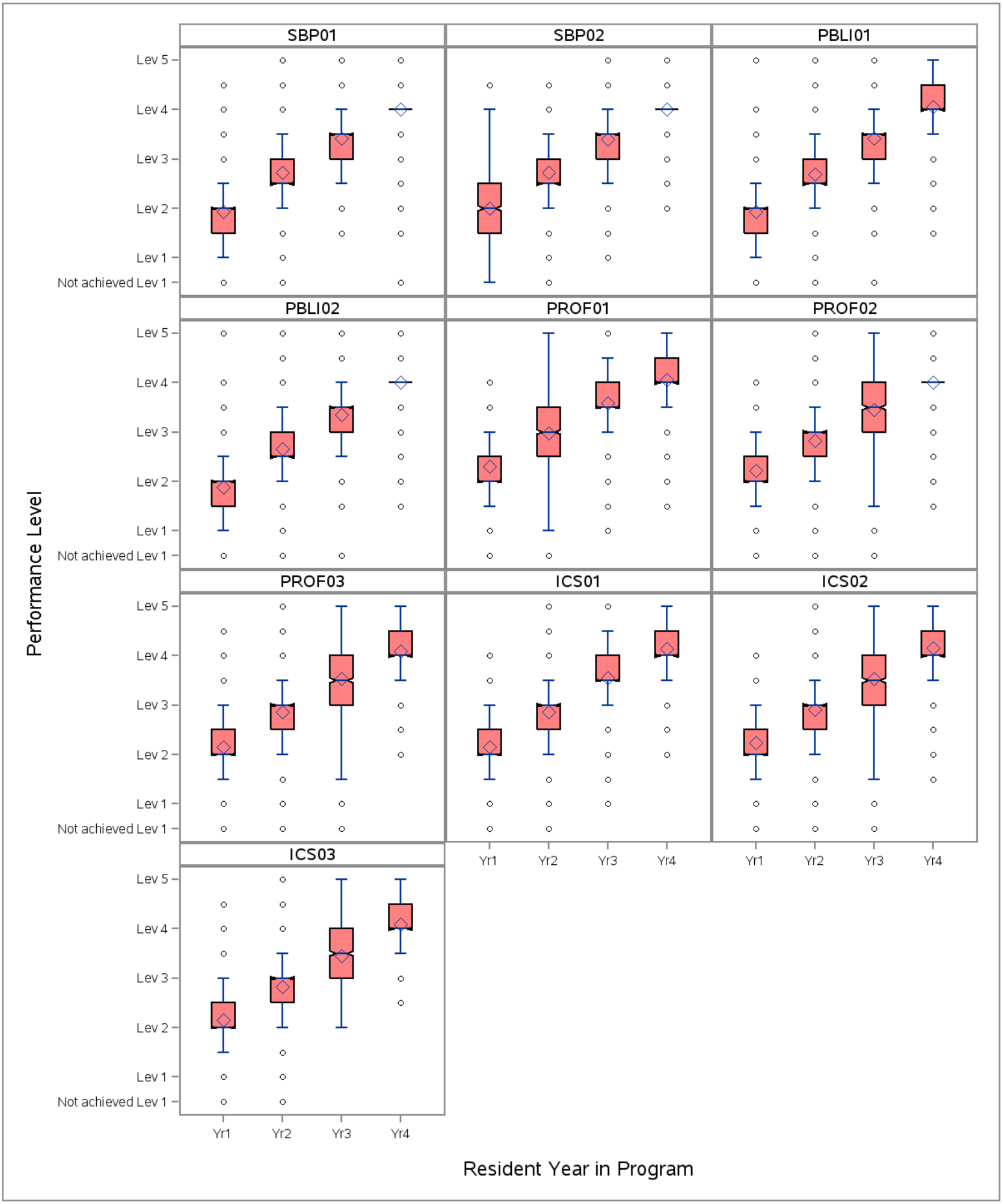


Table 16 - Ophthalmology (June 2016)

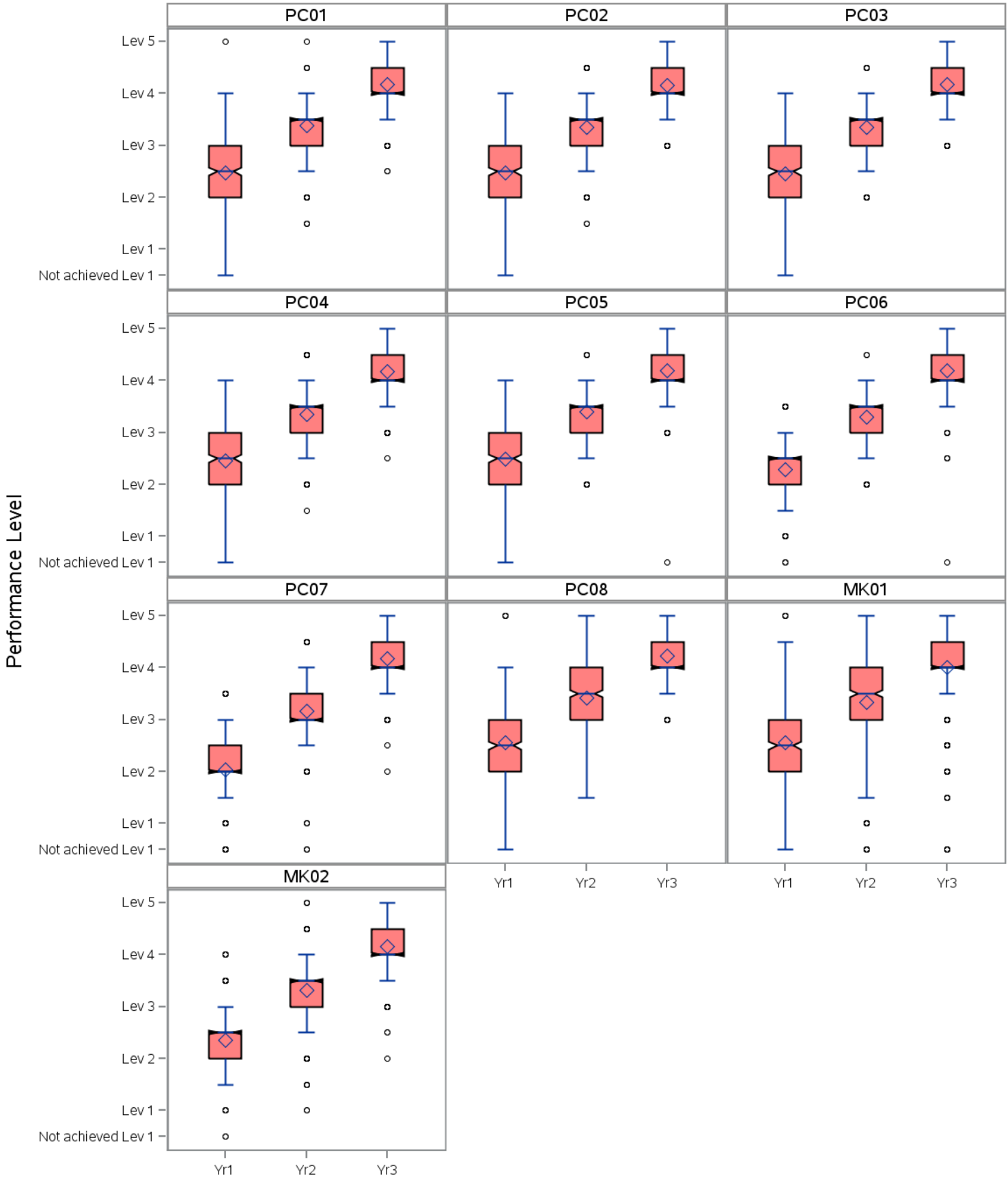


Table 16 - Ophthalmology (June 2016)

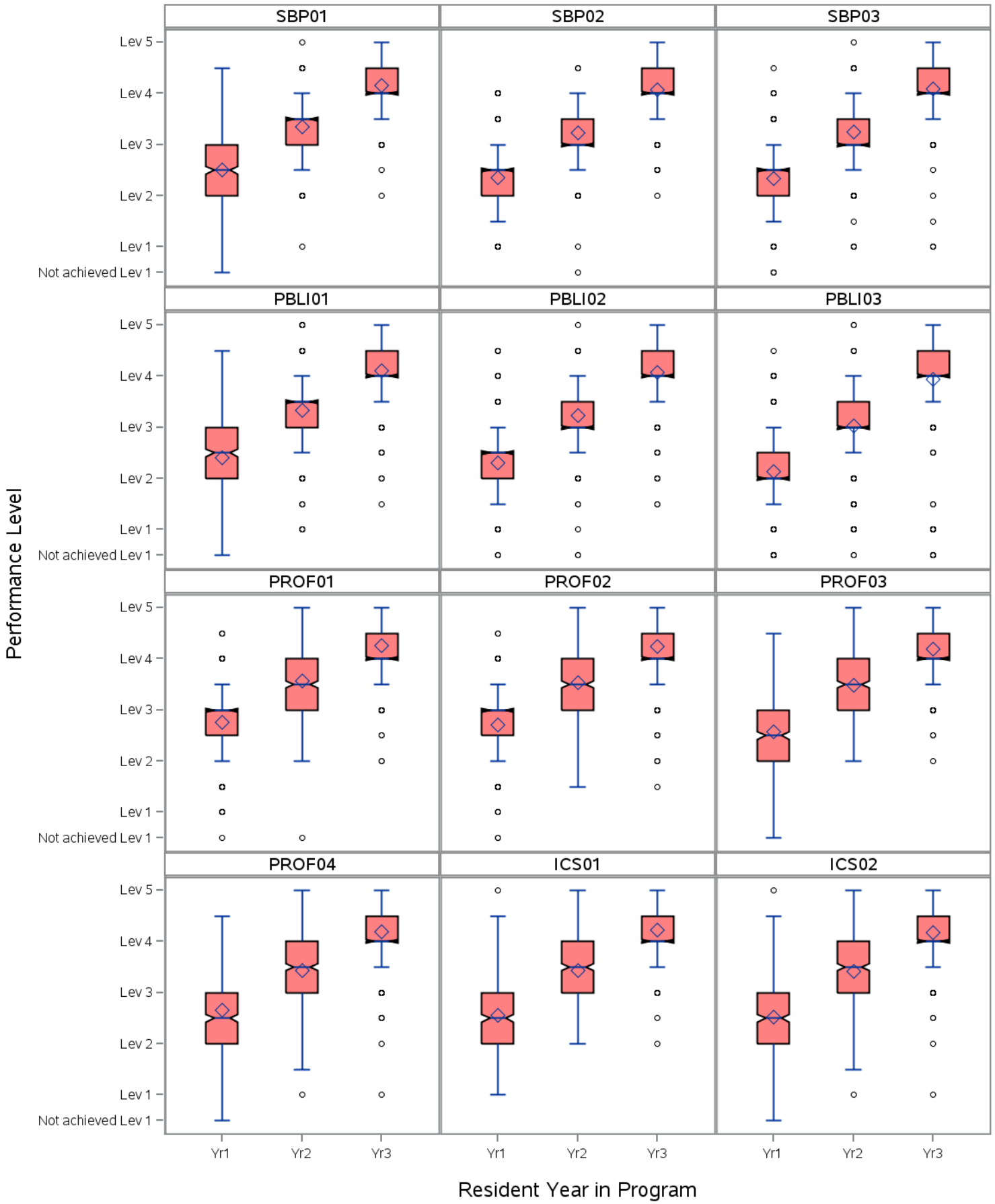


Table 16 - Ophthalmology (June 2016)

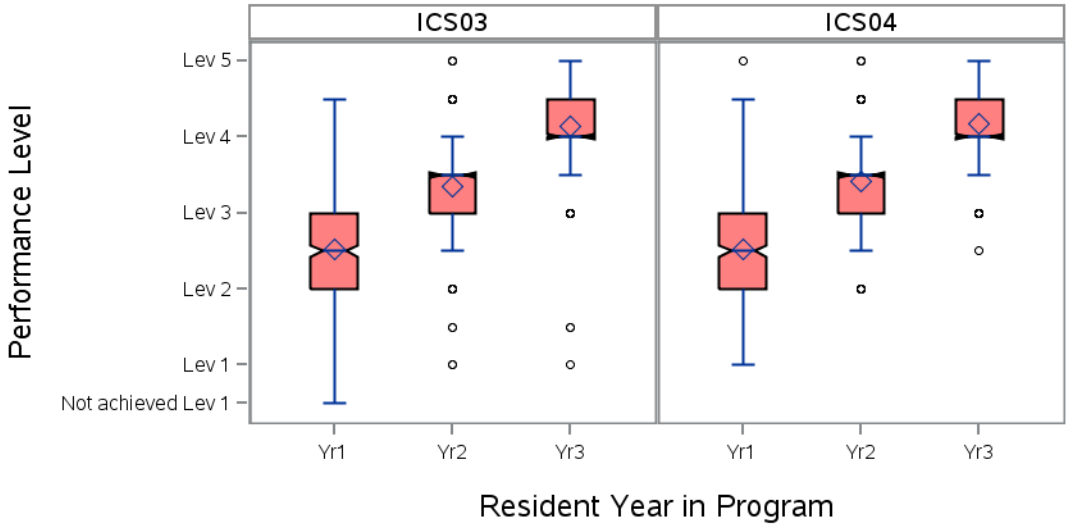


Table 17 - Orthopedic Surgery (June 2016)

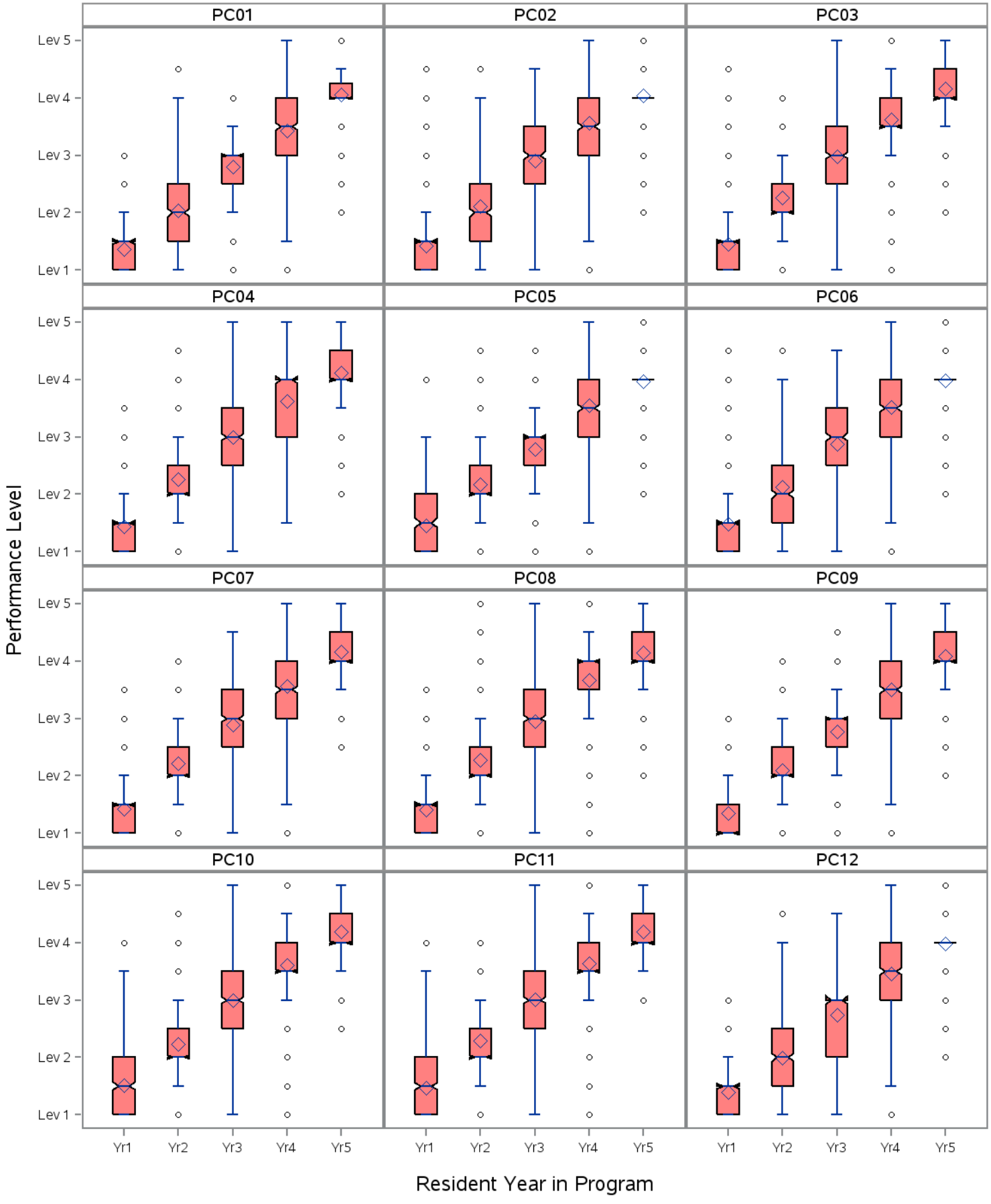


Table 17 - Orthopedic Surgery (June 2016)

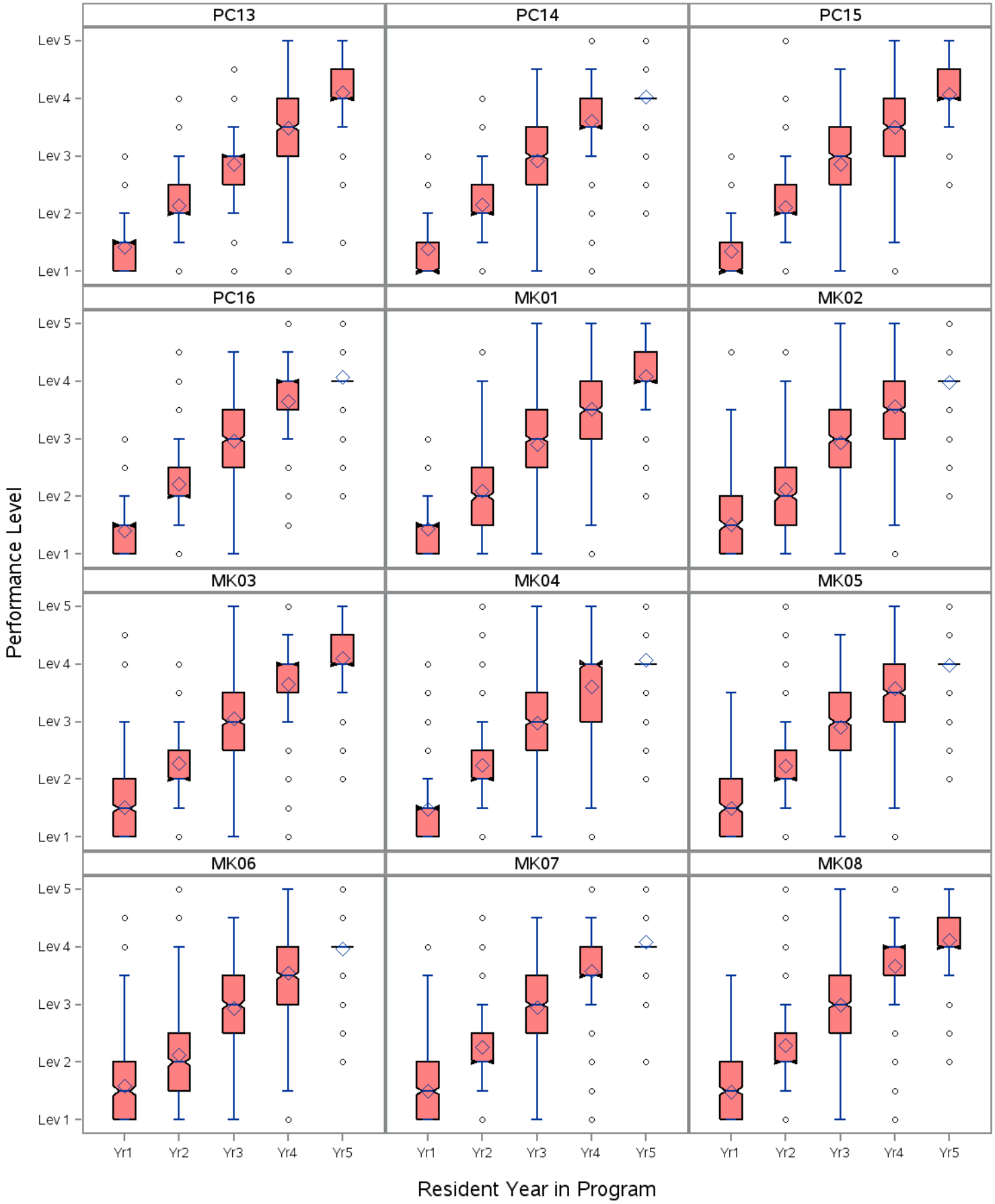


Table 17 - Orthopedic Surgery (June 2016)

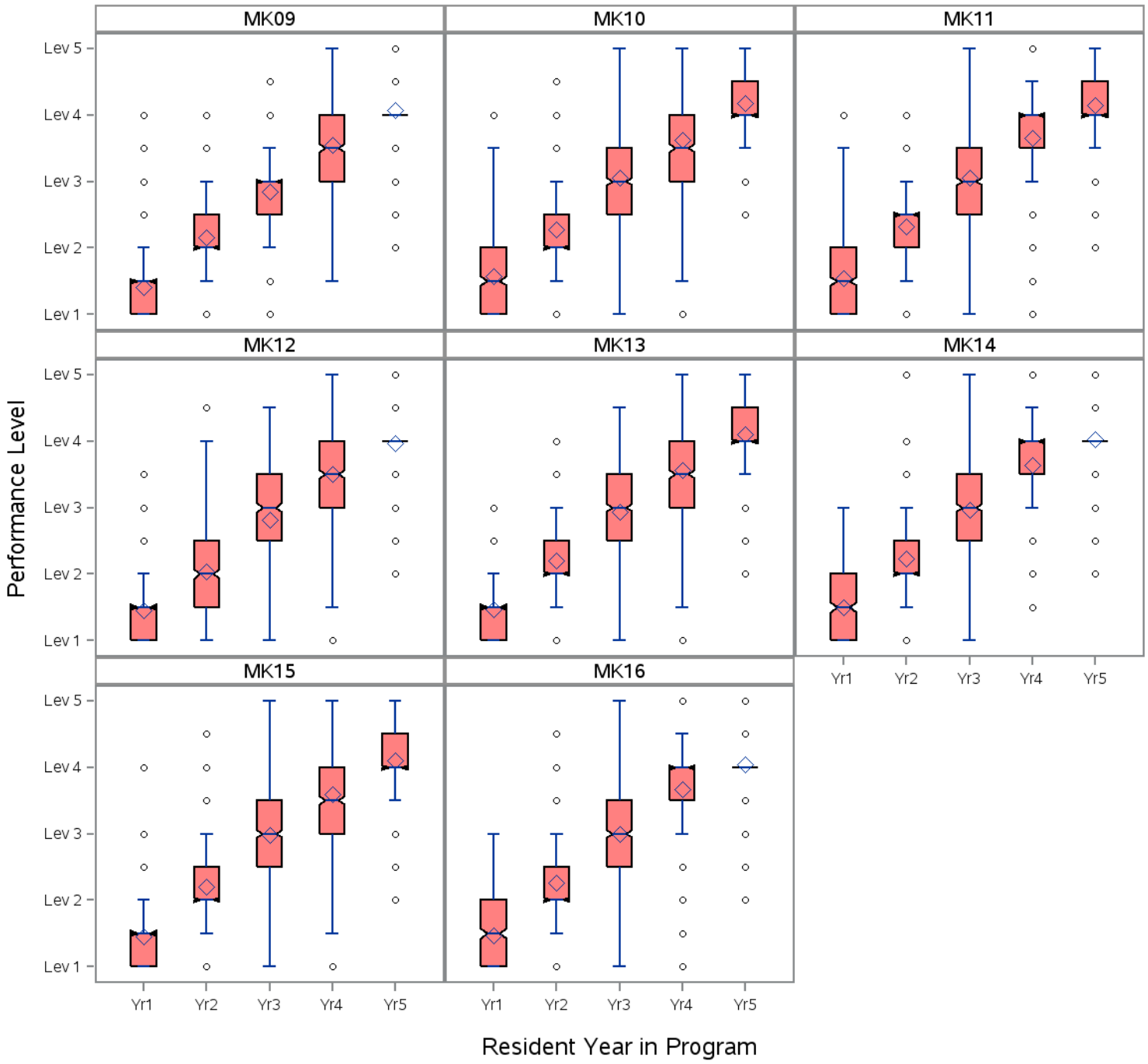


Table 17 - Orthopedic Surgery (June 2016)

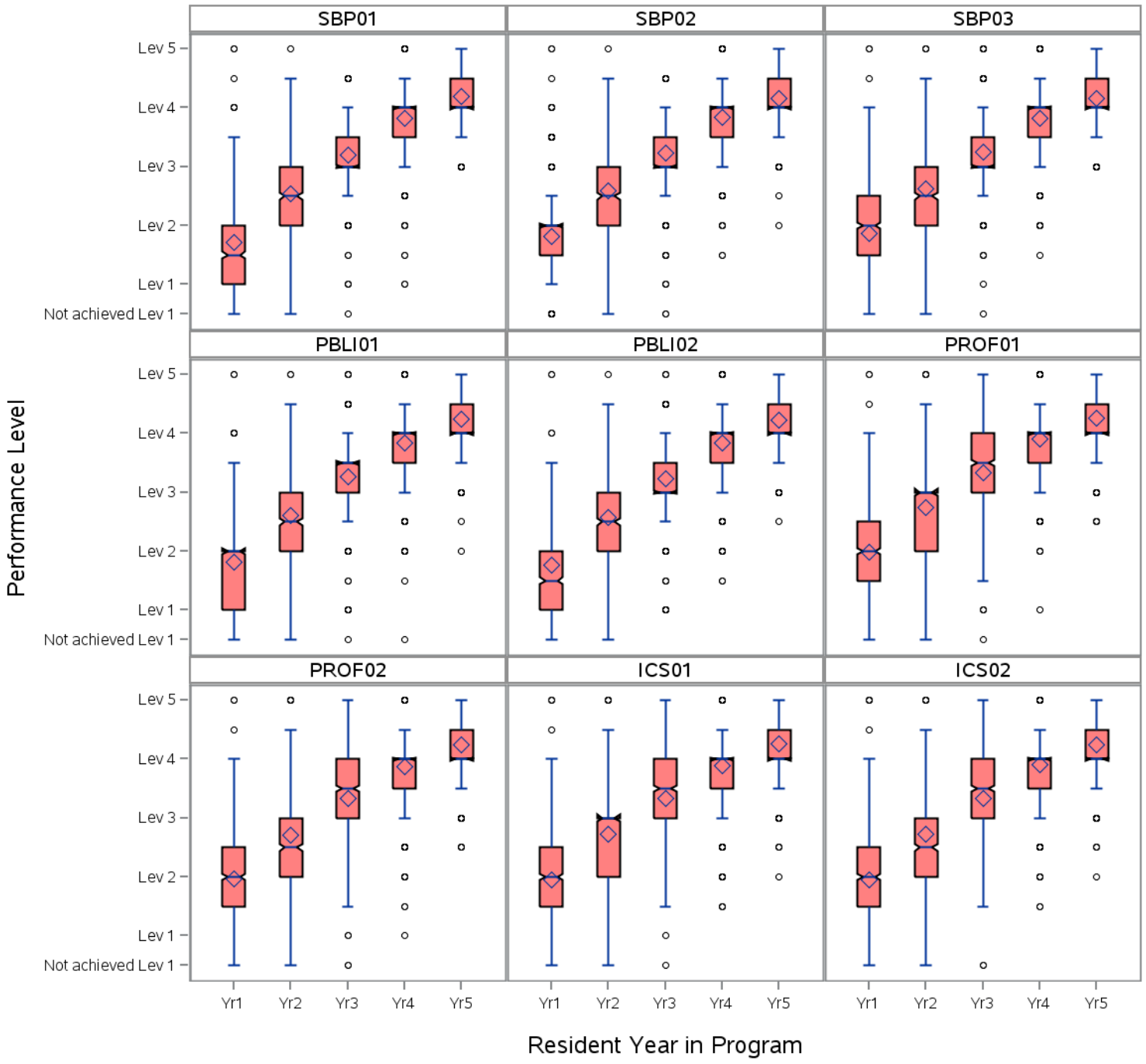


Table 18 - Otolaryngology (June 2016)

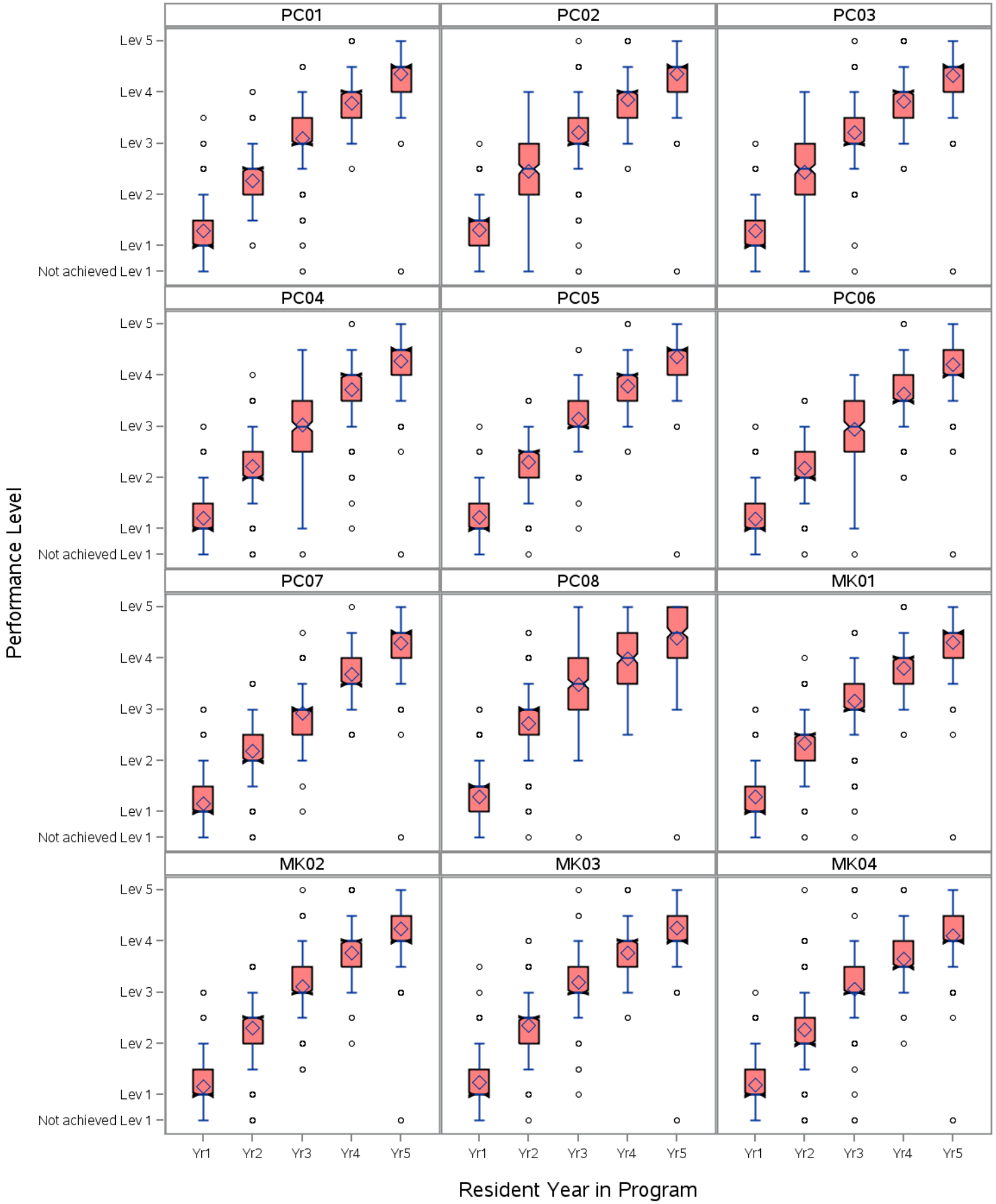


Table 18 - Otolaryngology (June 2016)

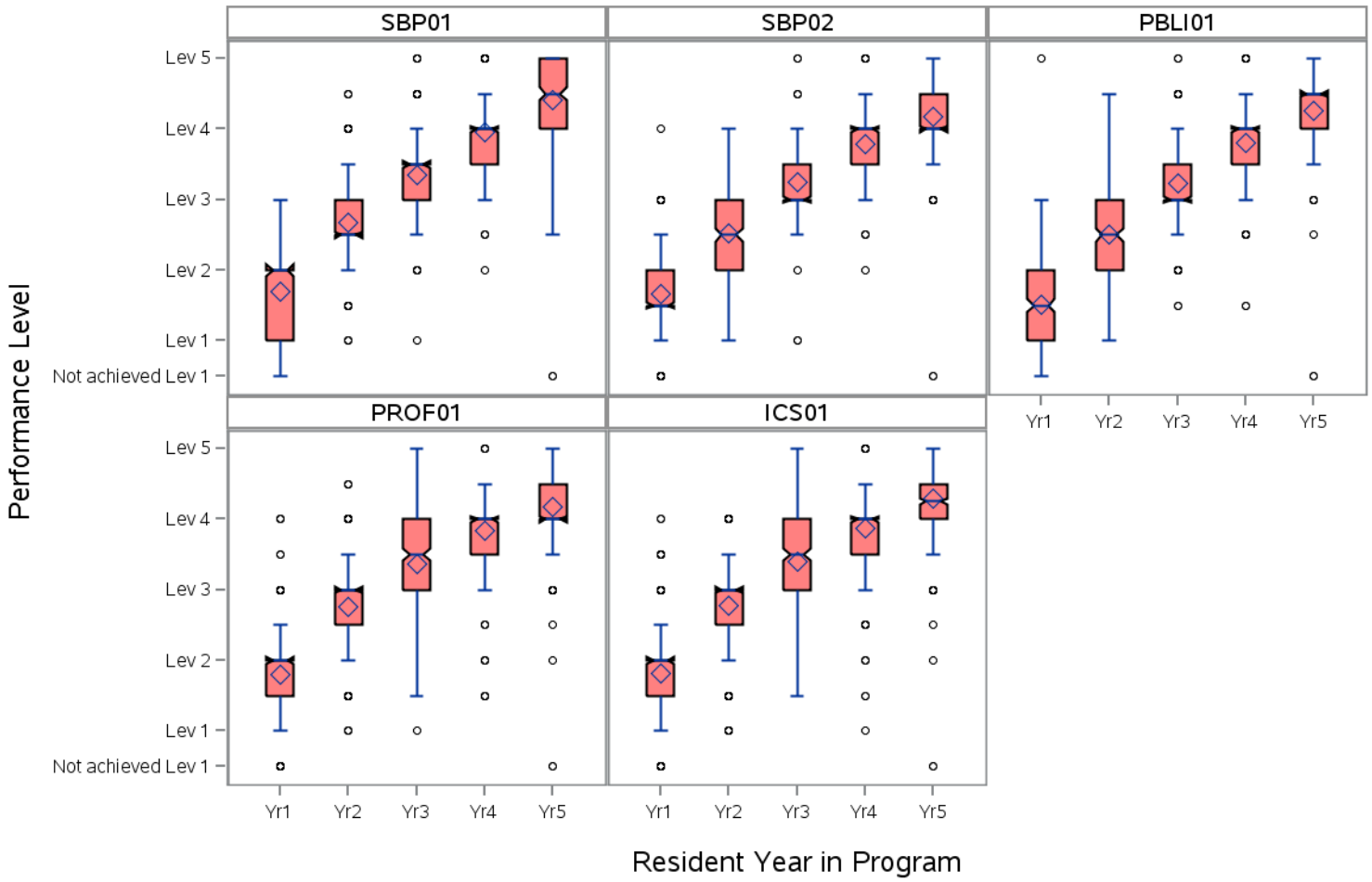


Table 19 - Pathology-Anatomic and Clinical (June 2016)

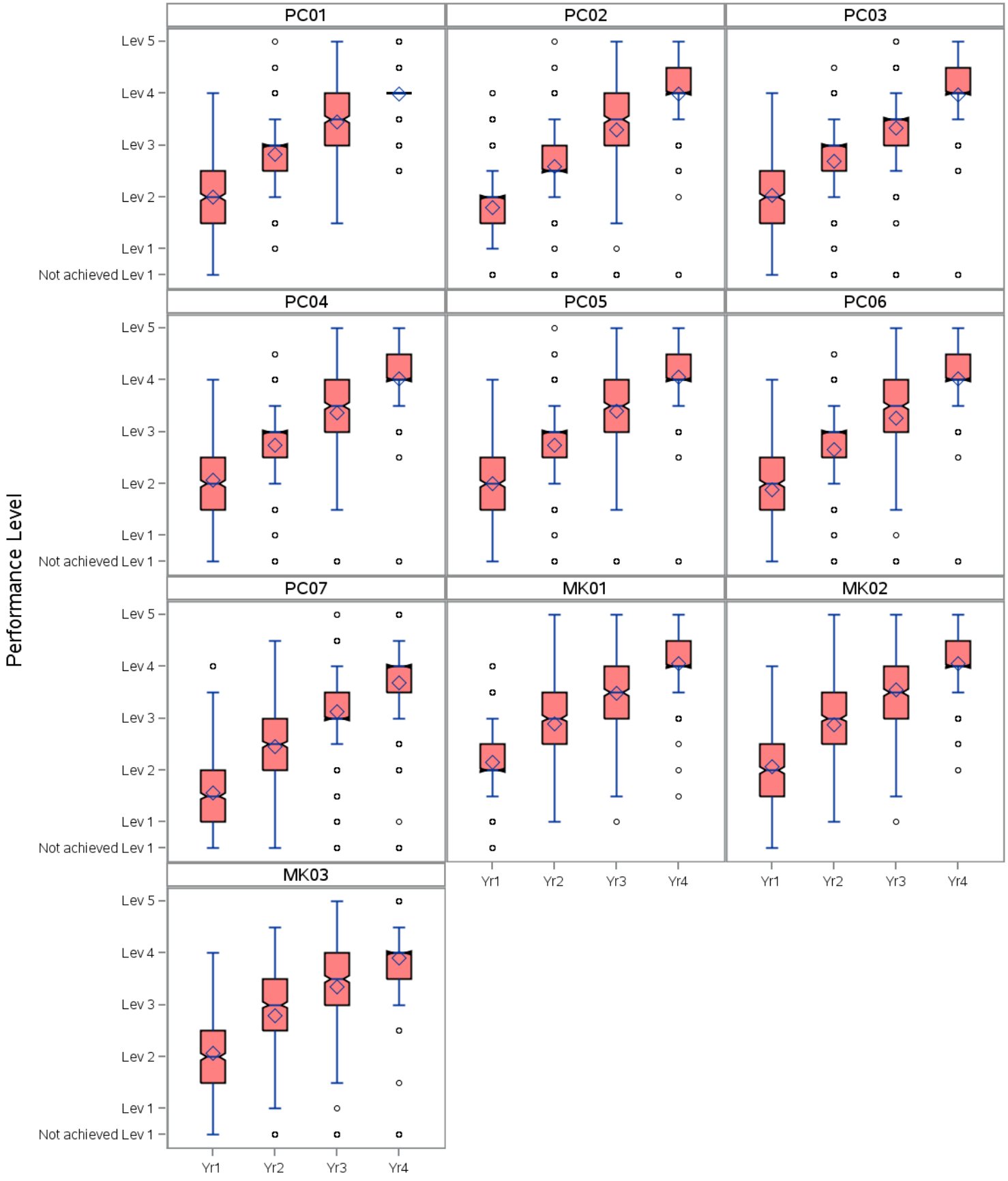


Table 19 - Pathology-Anatomic and Clinical (June 2016)

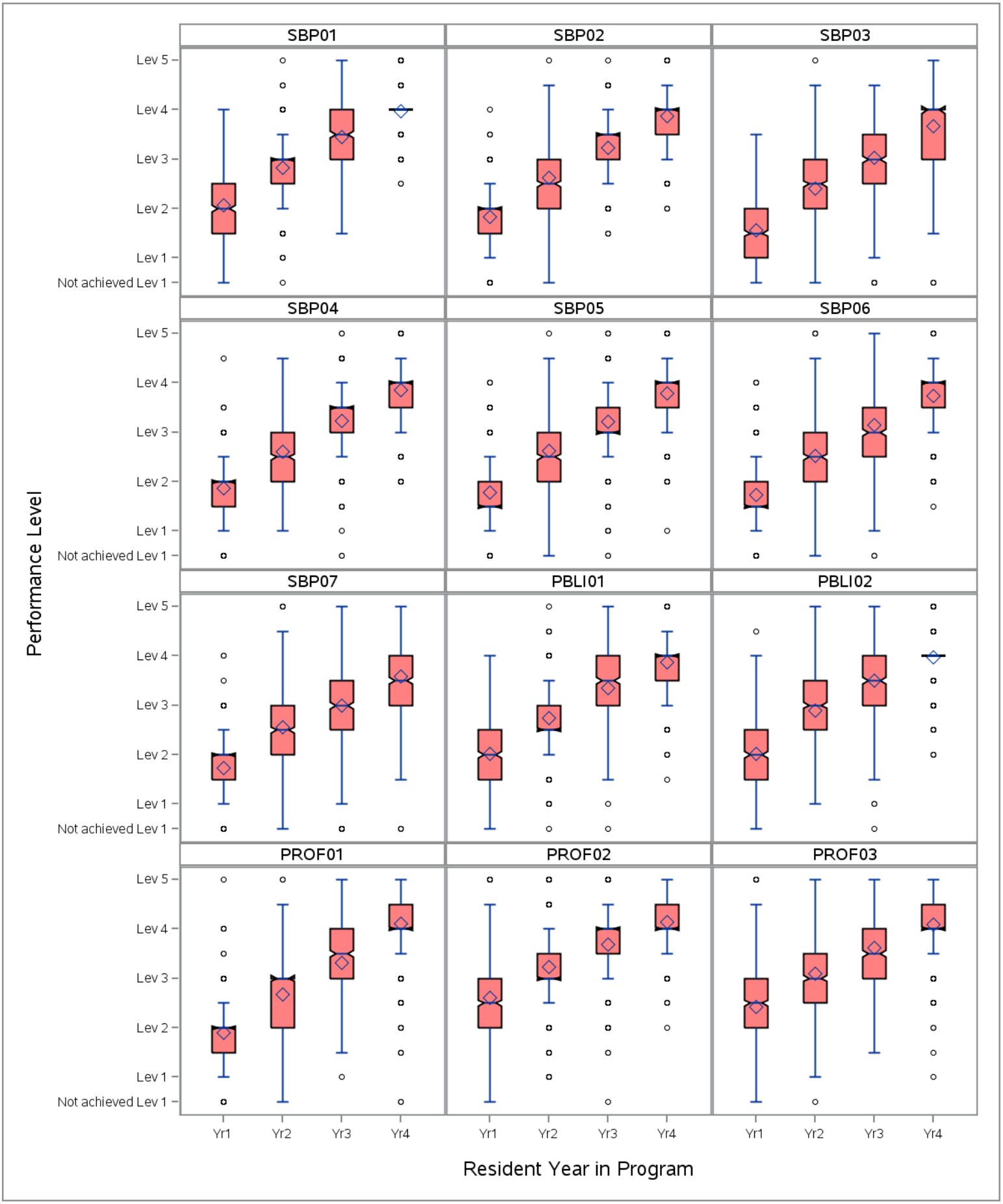


Table 19 - Pathology-Anatomic and Clinical (June 2016)

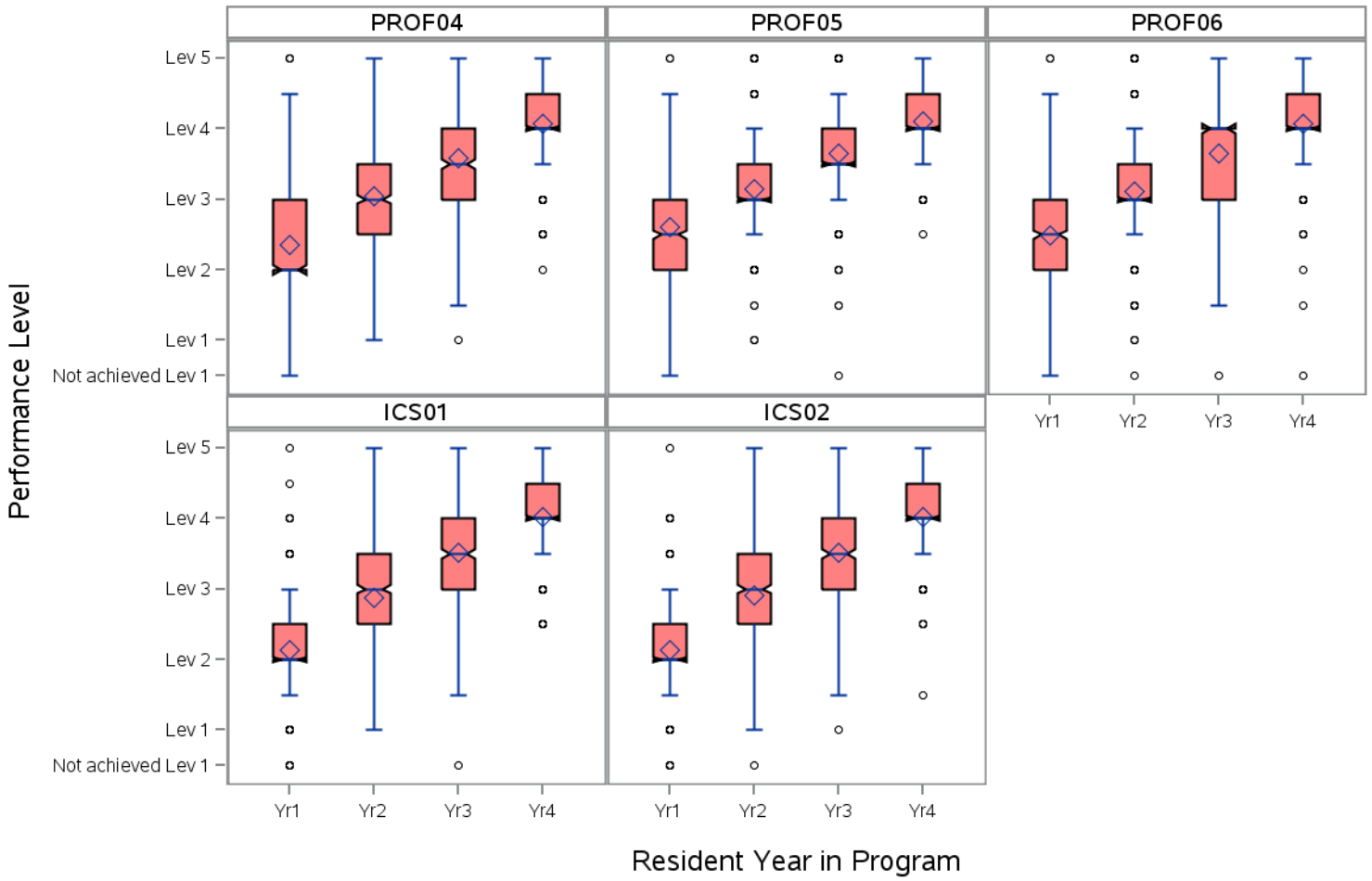


Table 20 - Pediatrics (June 2016)

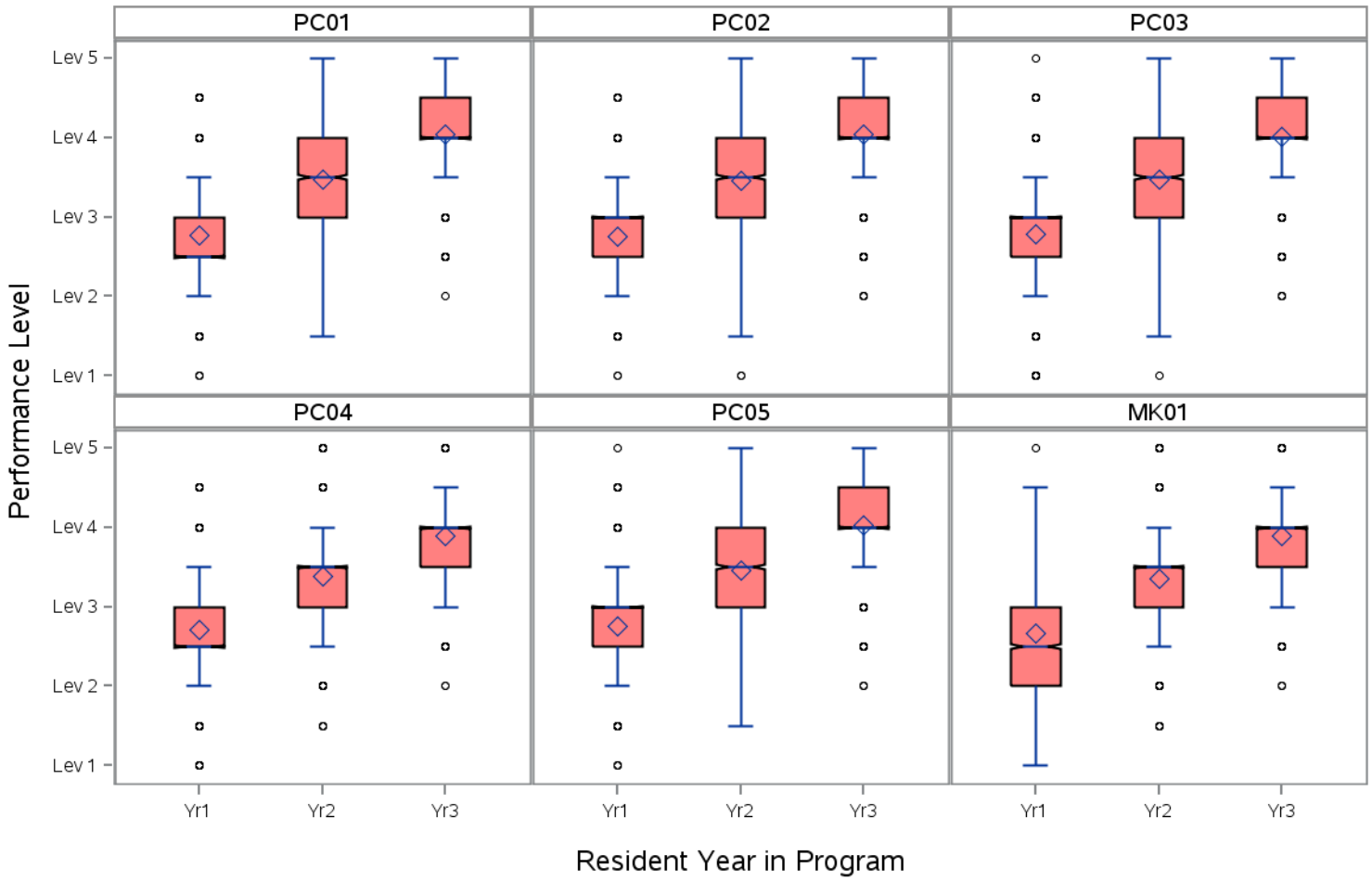


Table 20 - Pediatrics (June 2016)

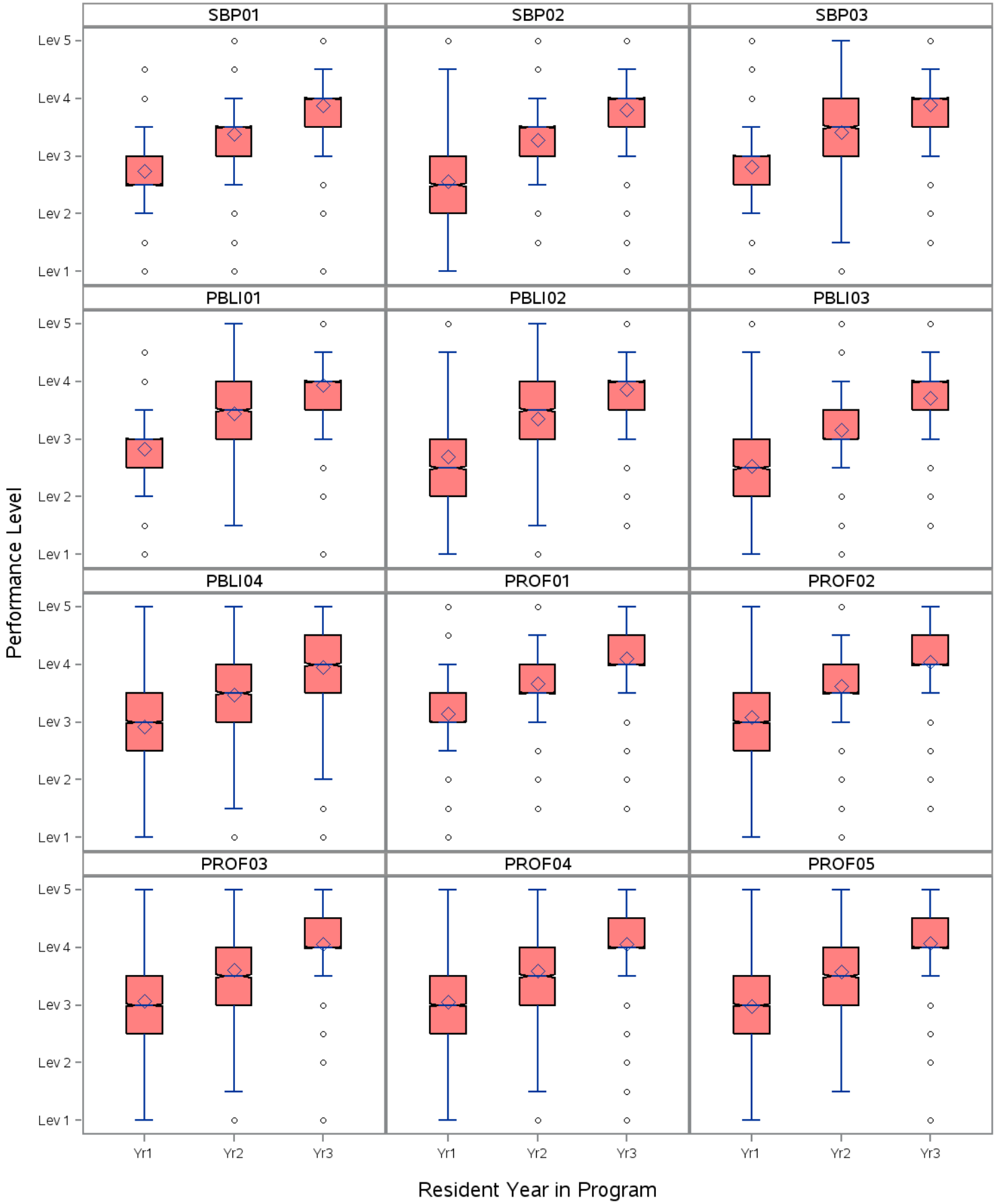


Table 20 - Pediatrics (June 2016)

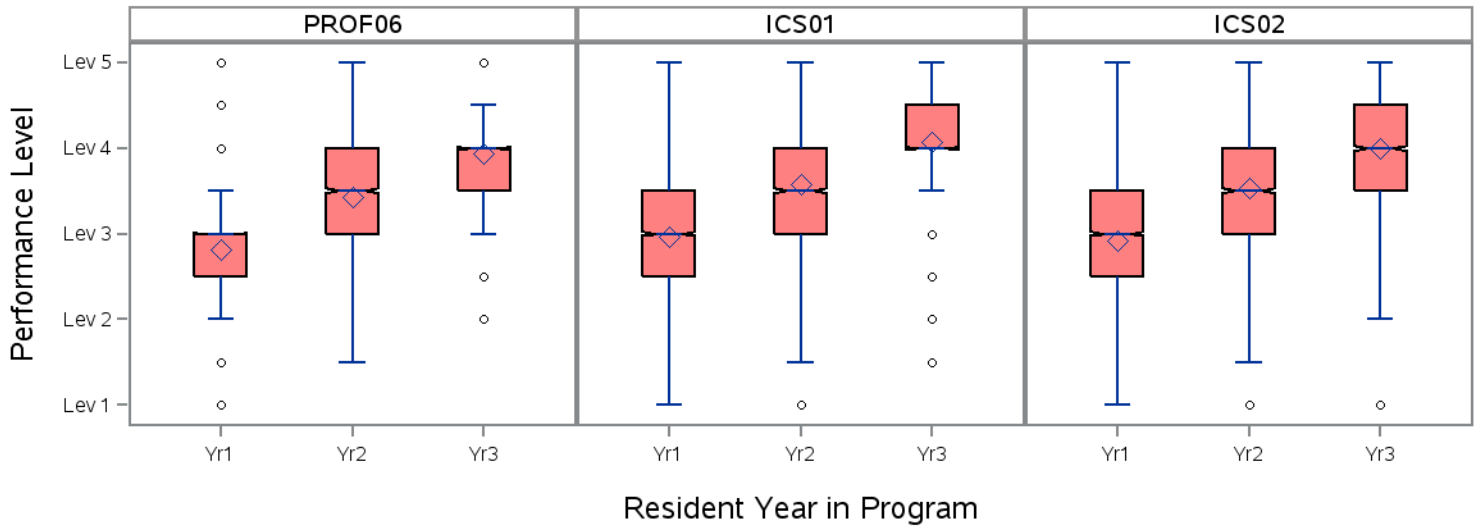


Table 21 - Physical Medicine and Rehabilitation (June 2016)

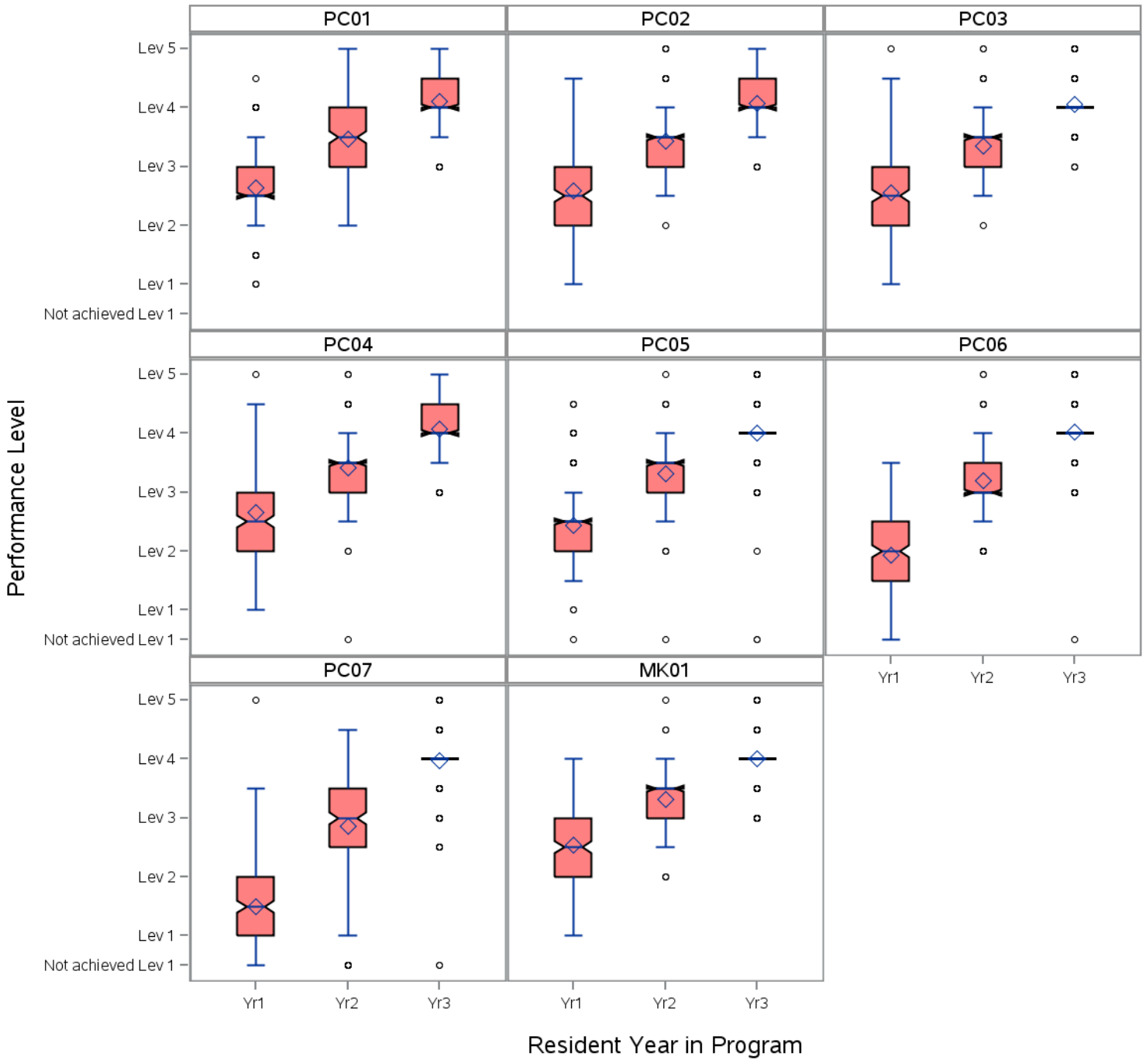


Table 21 - Physical Medicine and Rehabilitation (June 2016)

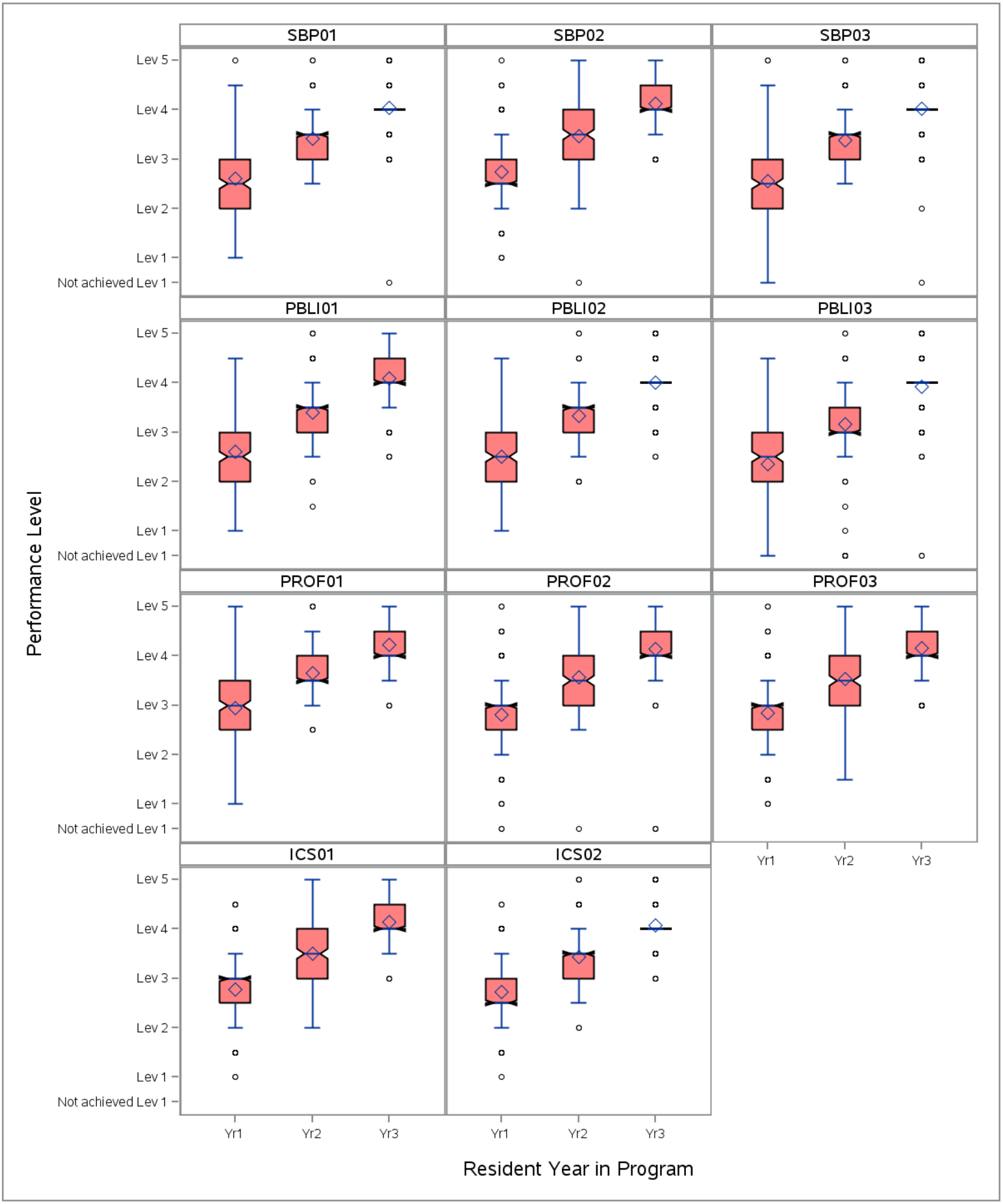


Table 22 - Plastic Surgery (June 2016)

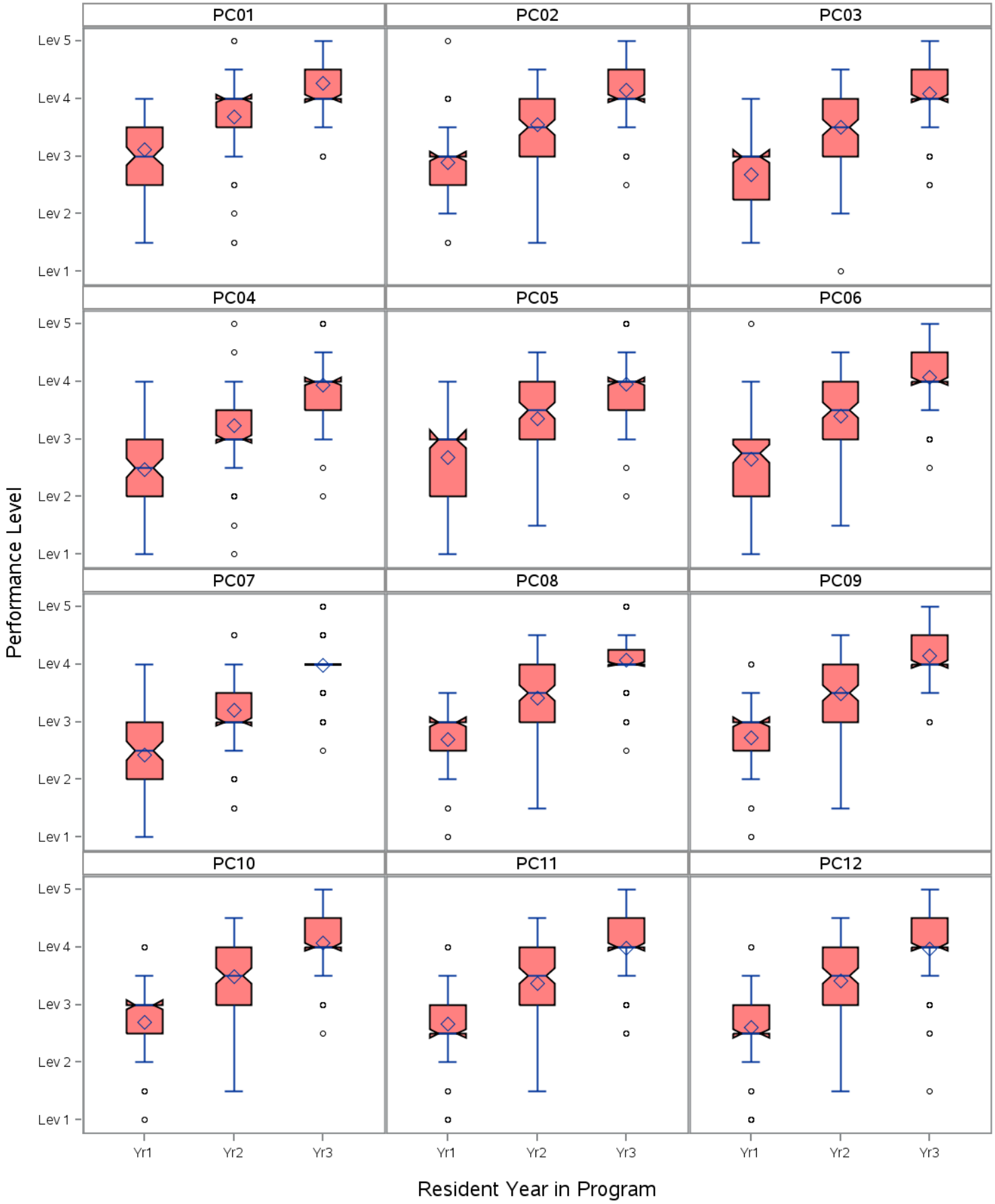


Table 22 - Plastic Surgery (June 2016)

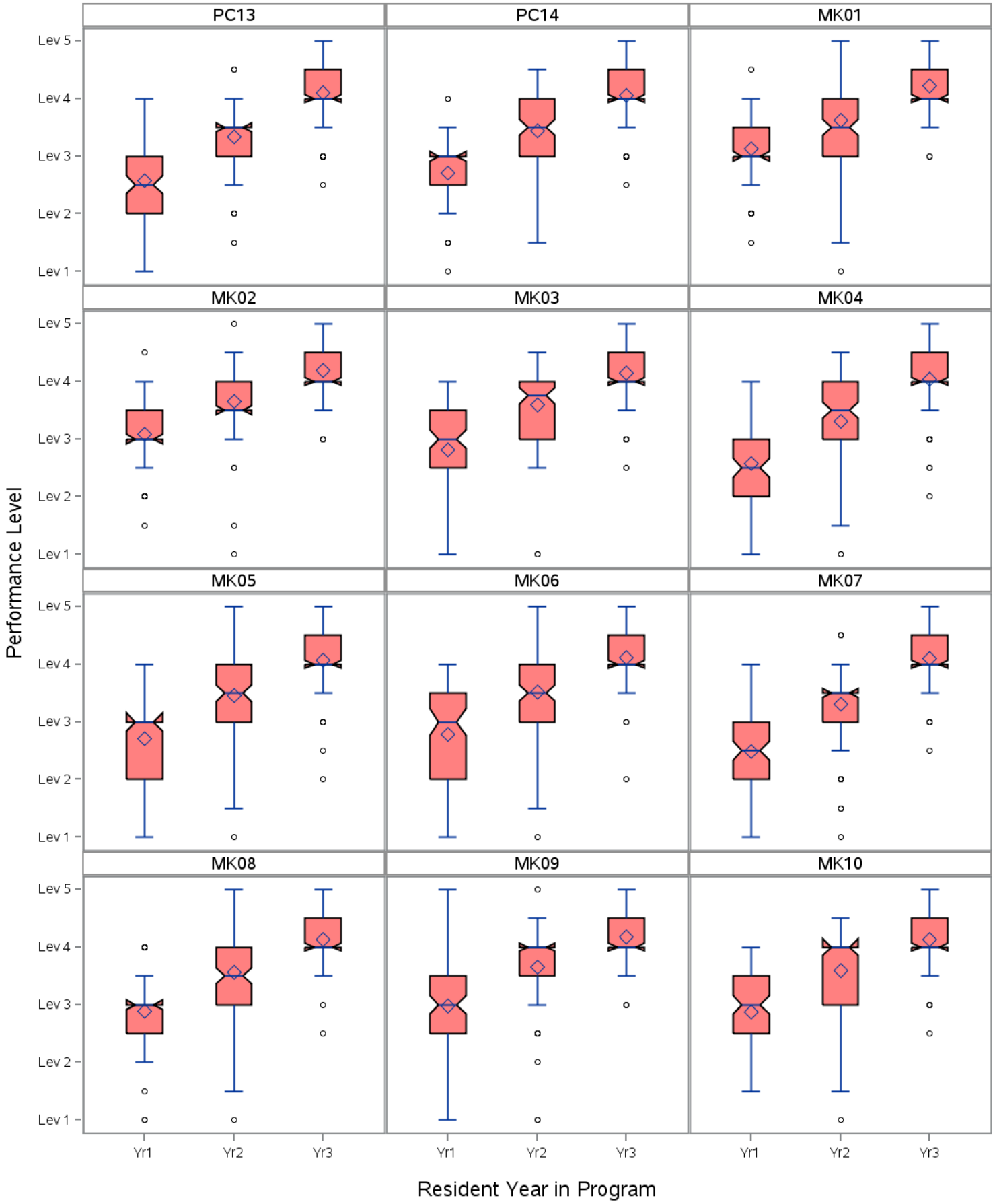


Table 22 - Plastic Surgery (June 2016)

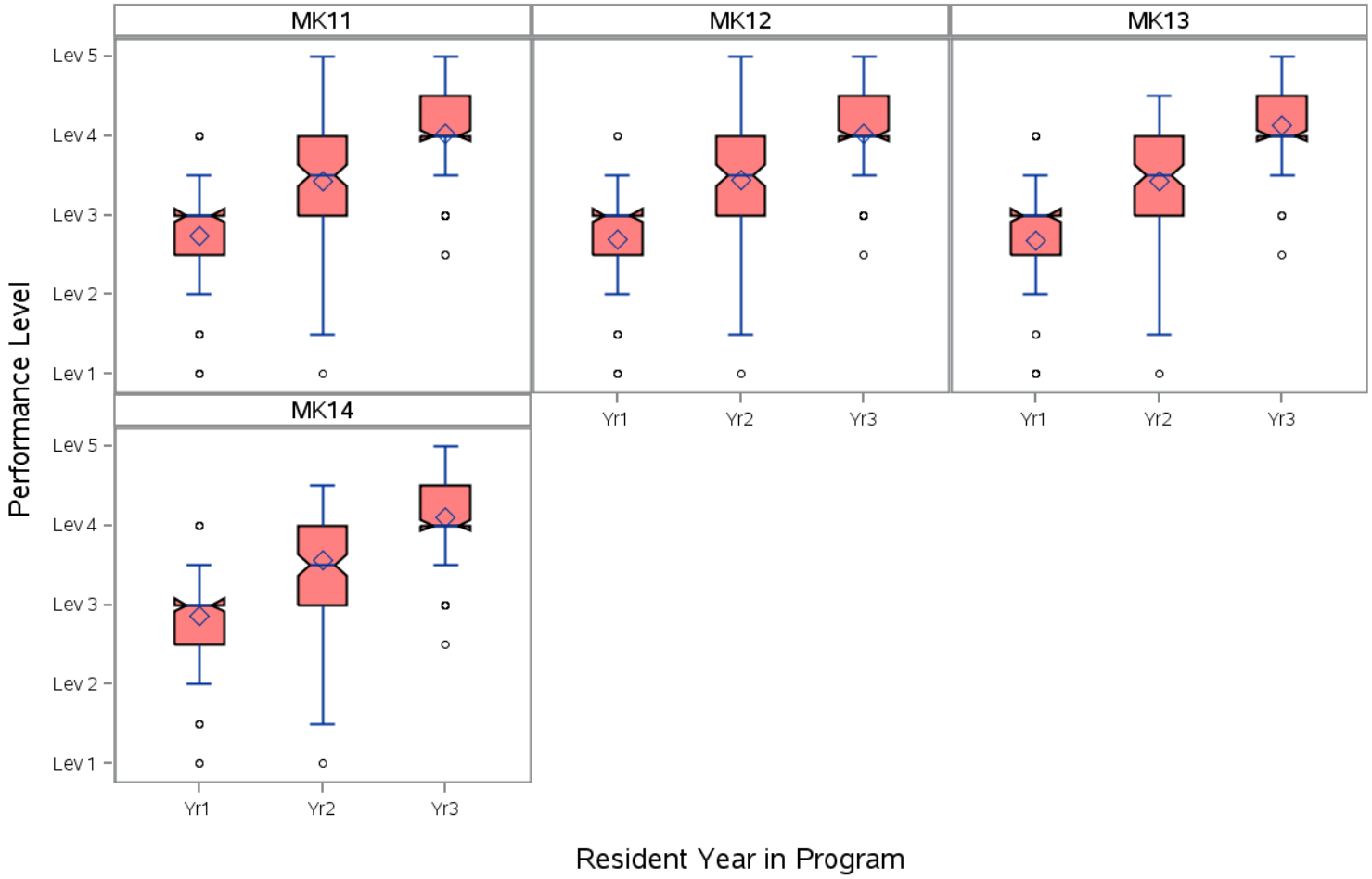


Table 22 - Plastic Surgery (June 2016)

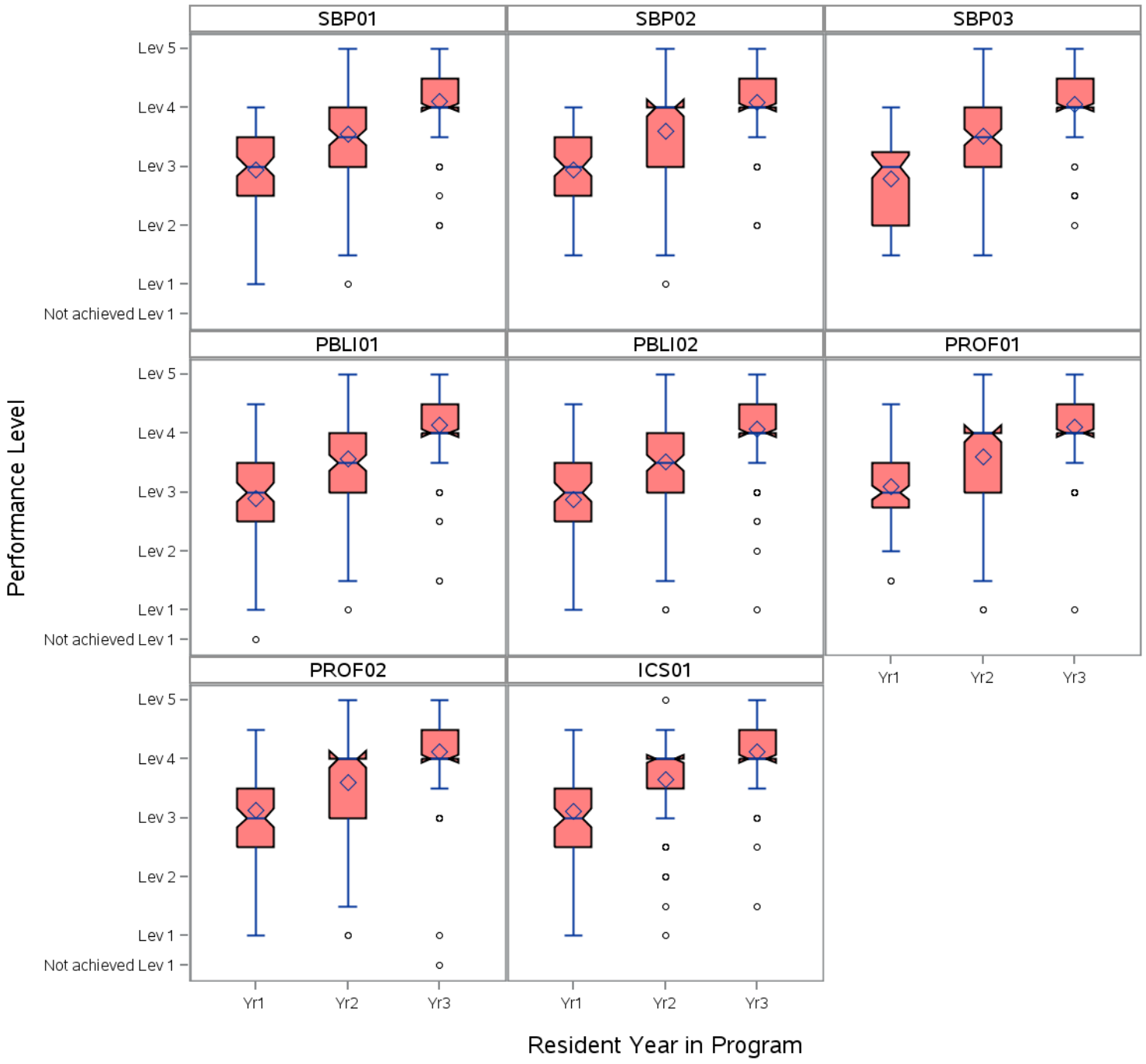


Table 23 - Plastic Surgery - Integrated (June 2016)

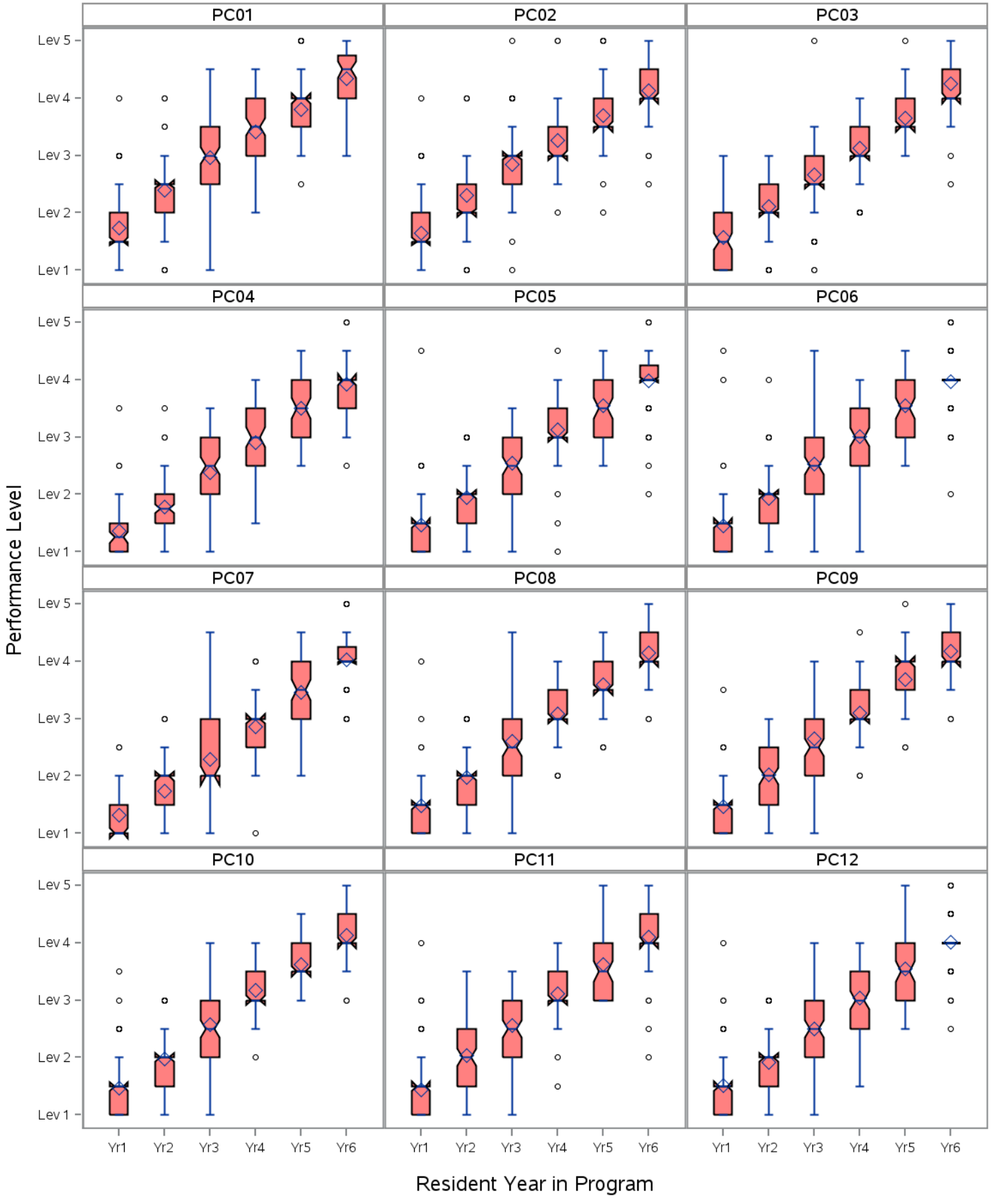


Table 23 - Plastic Surgery - Integrated (June 2016)

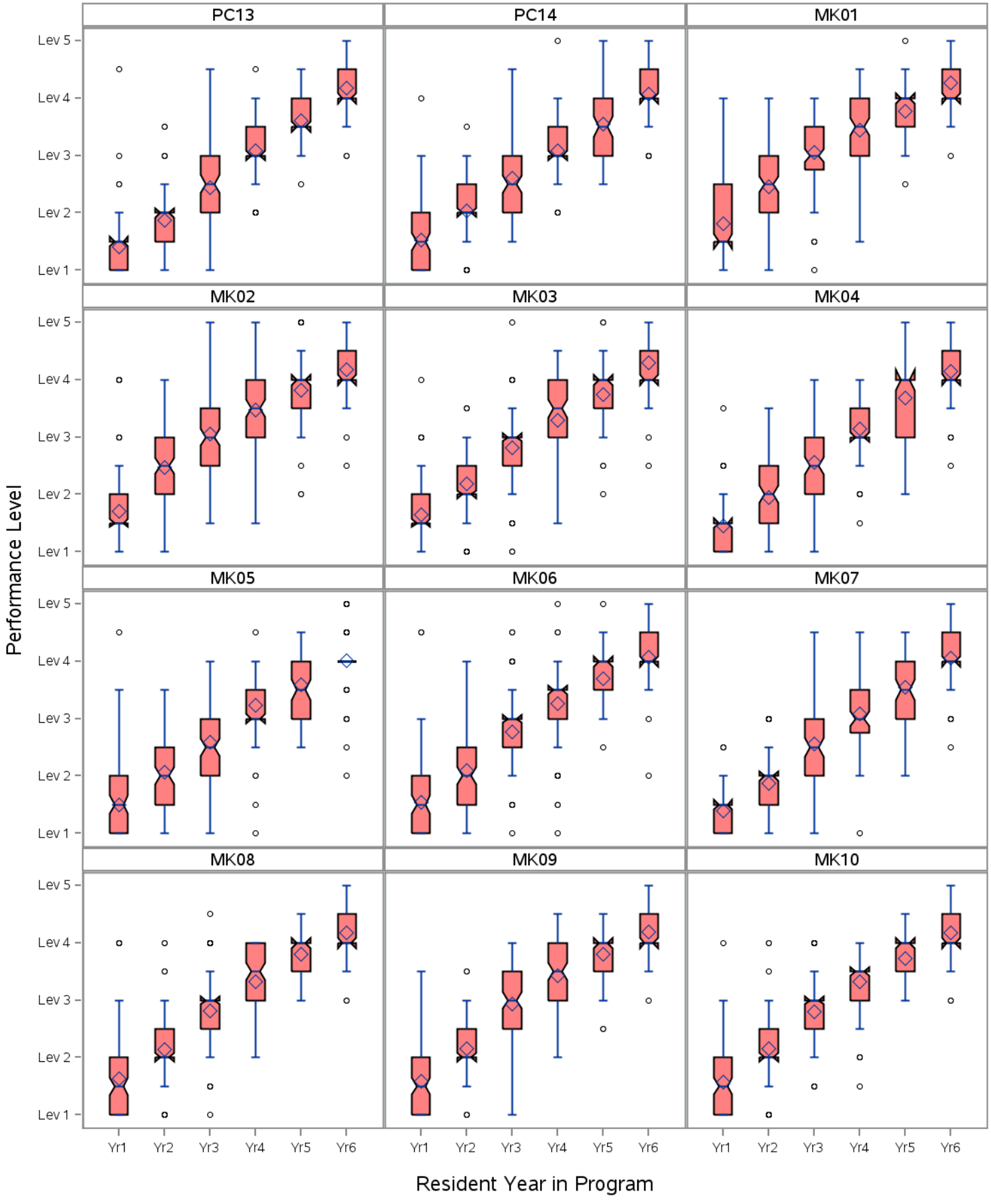


Table 23 - Plastic Surgery - Integrated (June 2016)

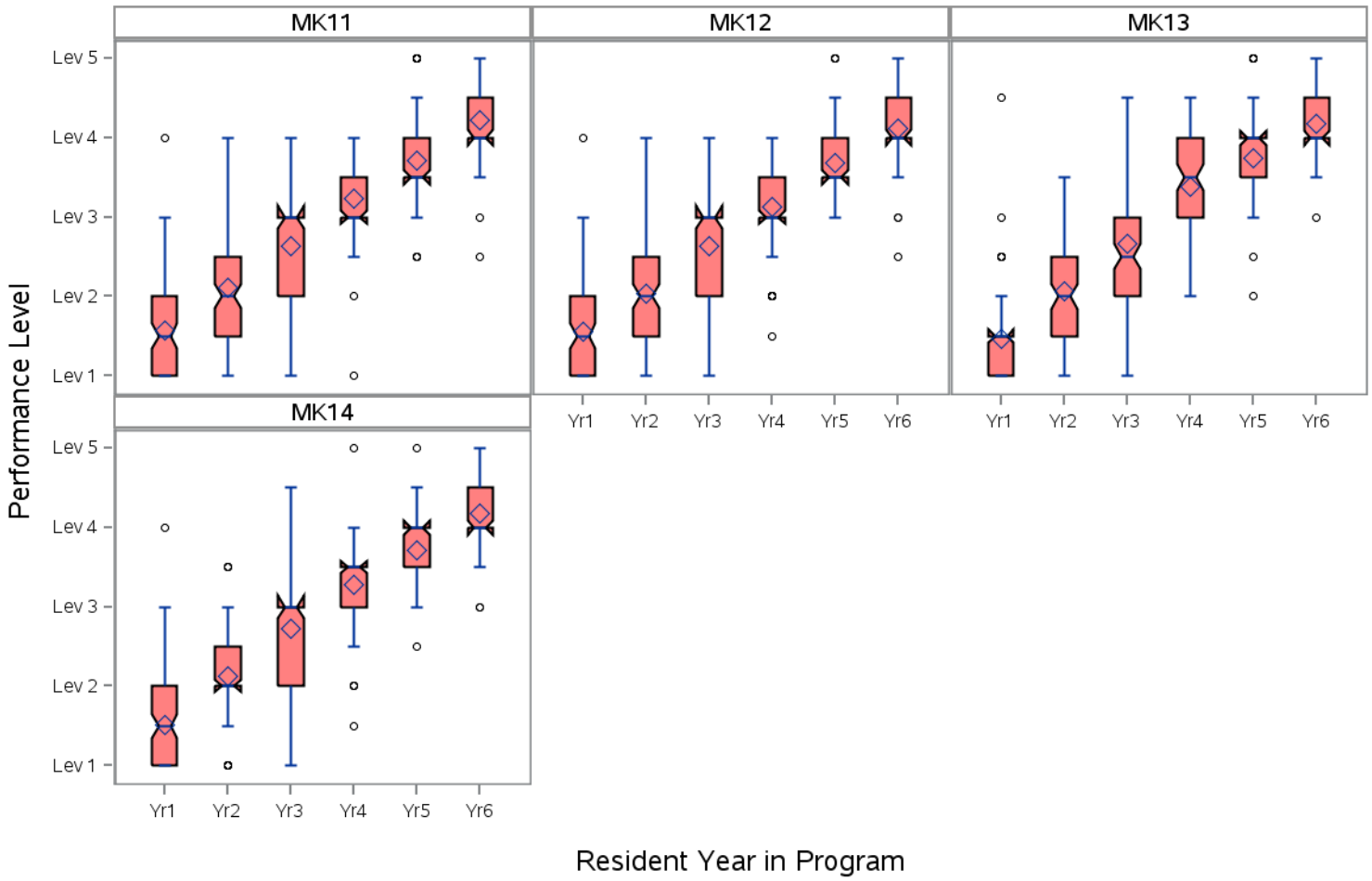


Table 23 - Plastic Surgery - Integrated (June 2016)

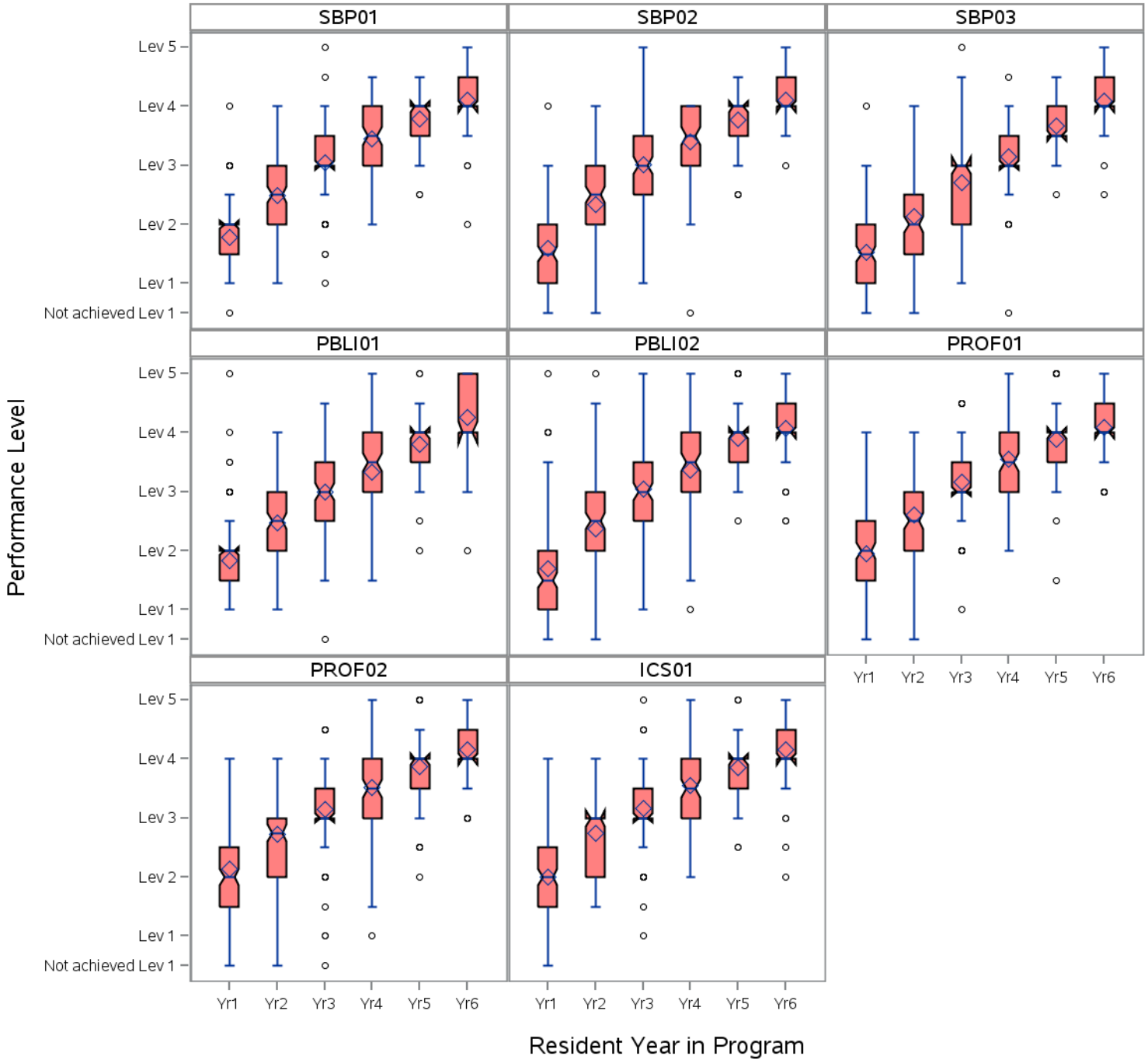


Table 24 - PM_Aerospace Medicine (June 2016)

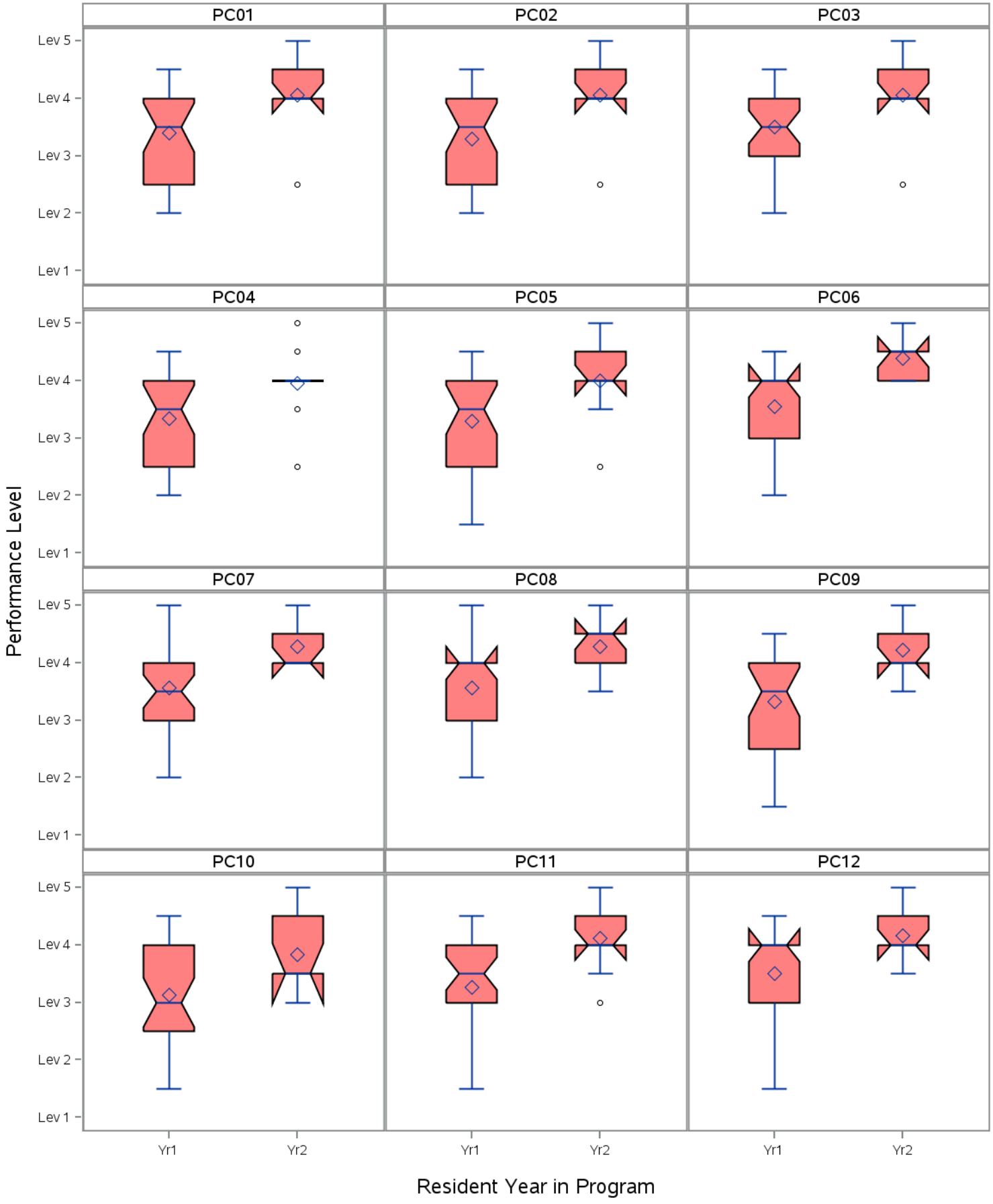


Table 24 - PM_Aerospace Medicine (June 2016)

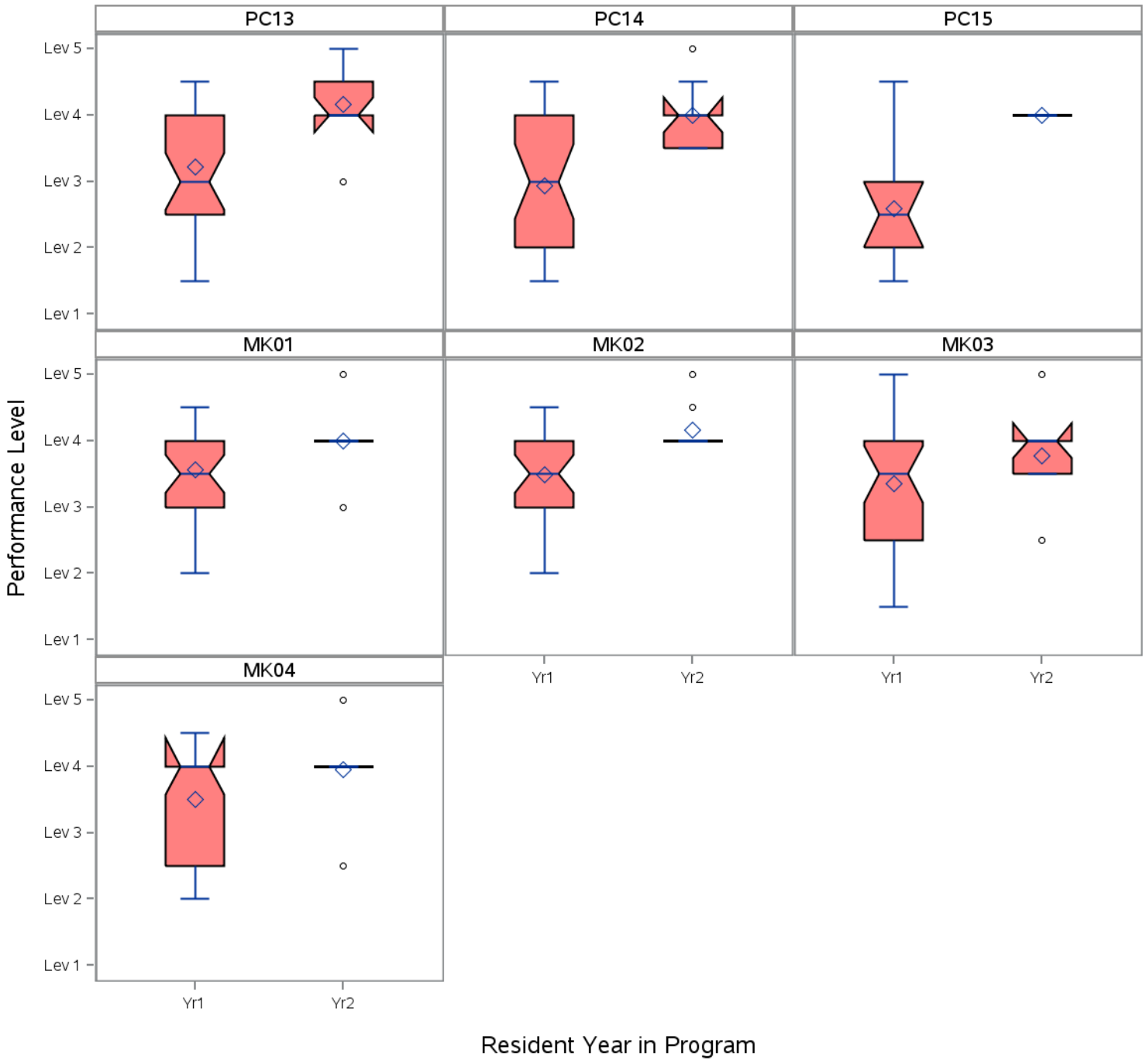


Table 24 - PM_Aerospace Medicine (June 2016)

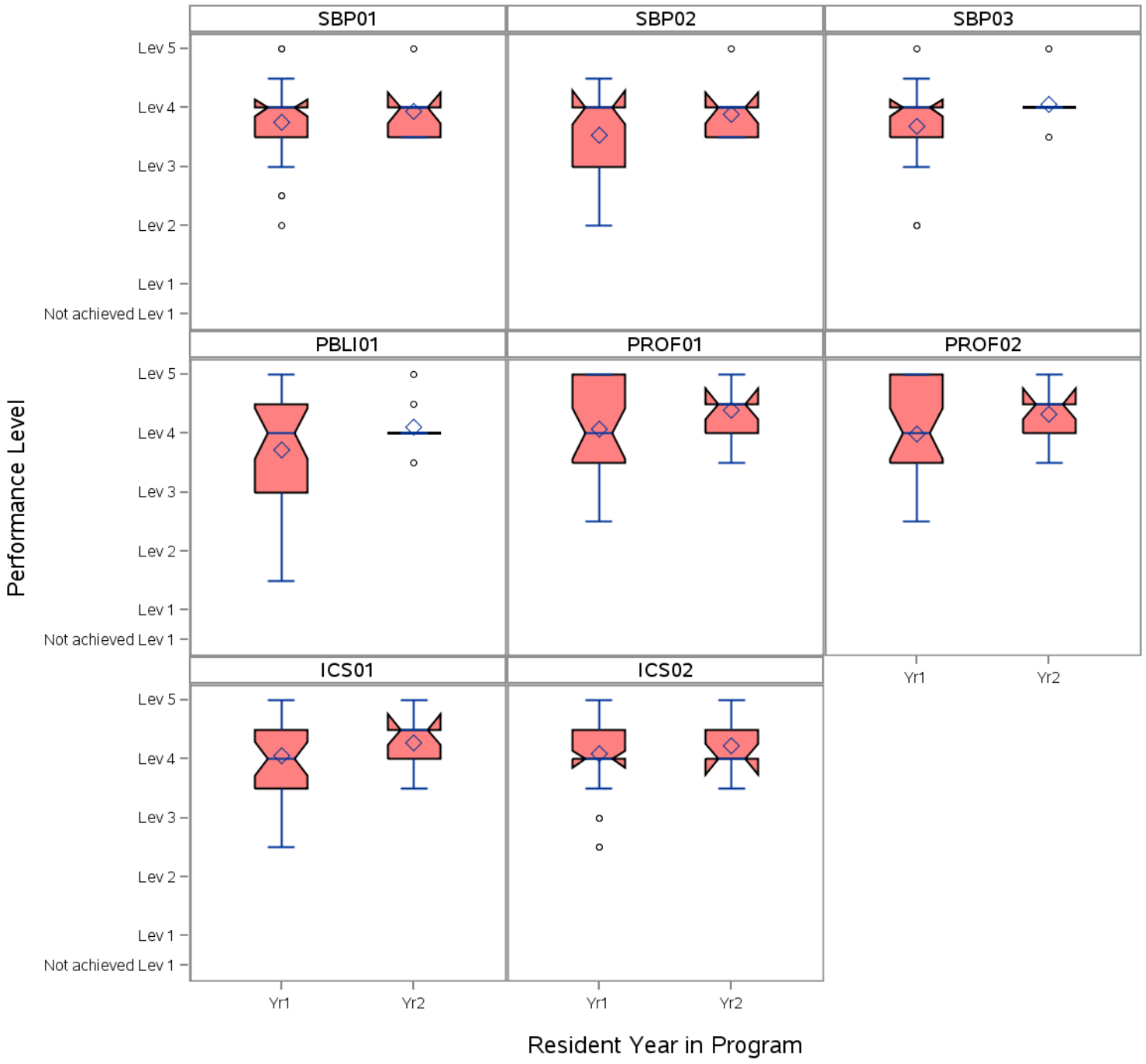


Table 25 - PM_Occupational Medicine (June 2016)

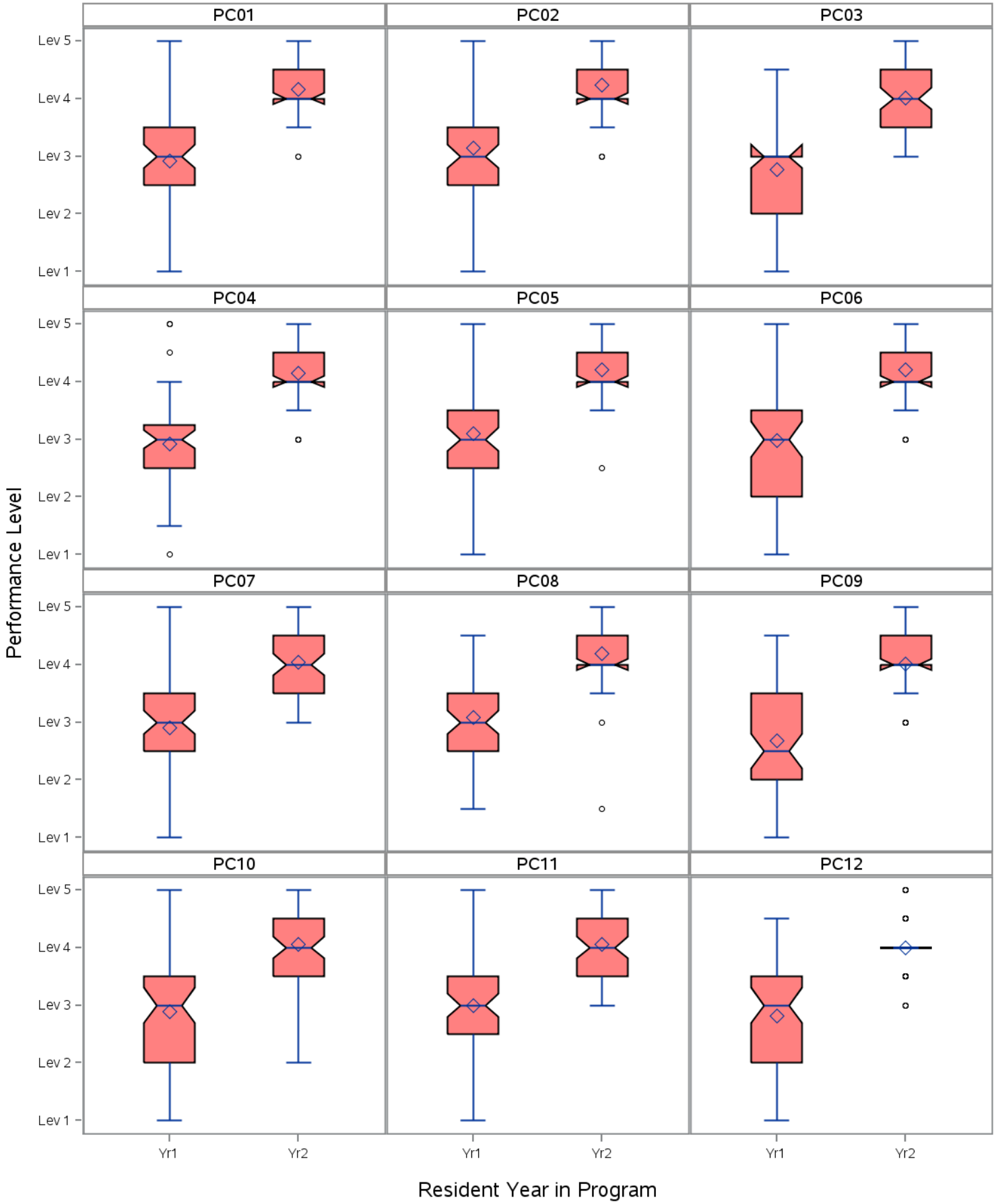


Table 25 - PM_Occupational Medicine (June 2016)

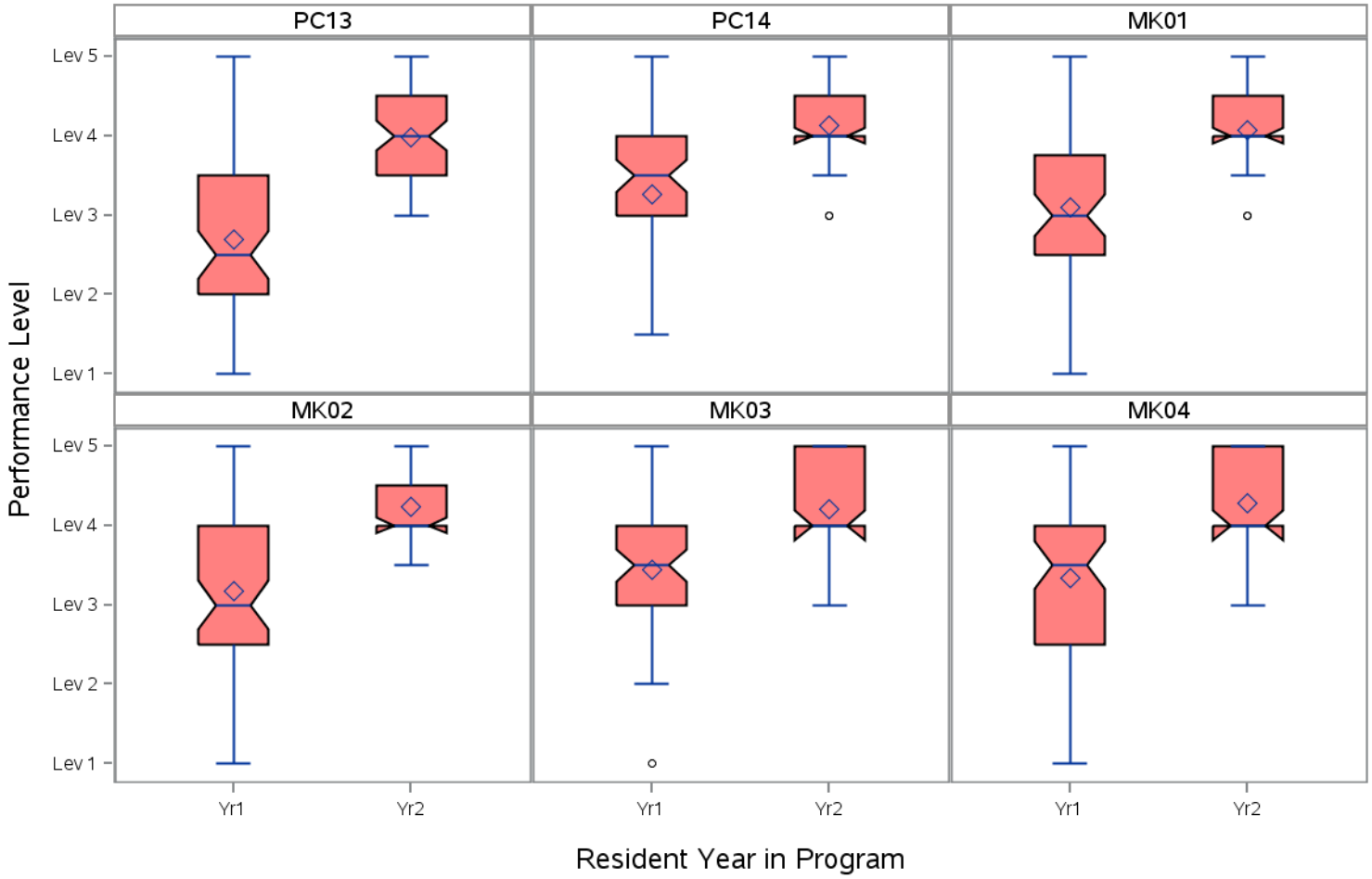


Table 25 - PM_Occupational Medicine (June 2016)

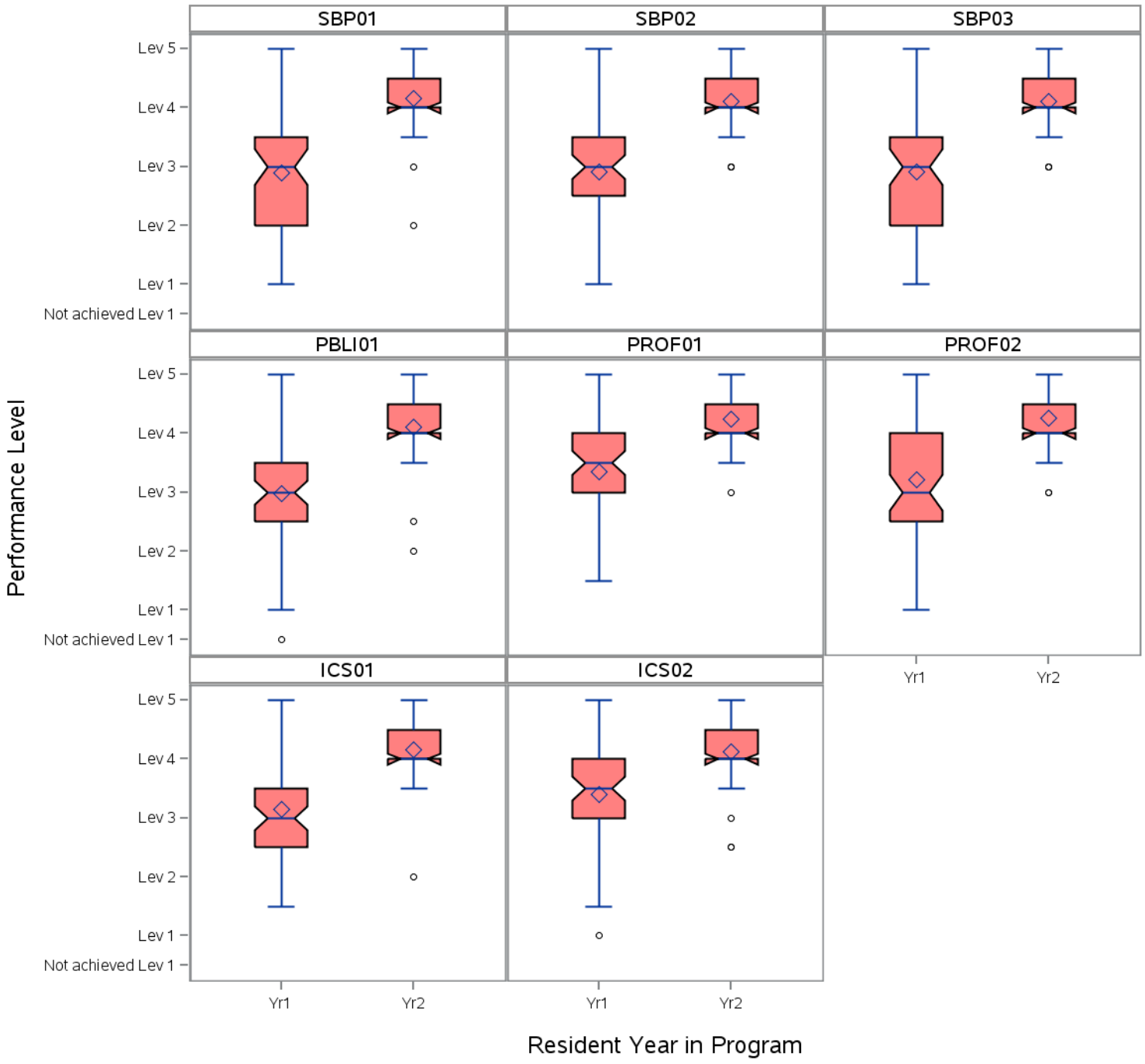


Table 26 - PM_Public Health and General Preventive Medicine (June 2016)

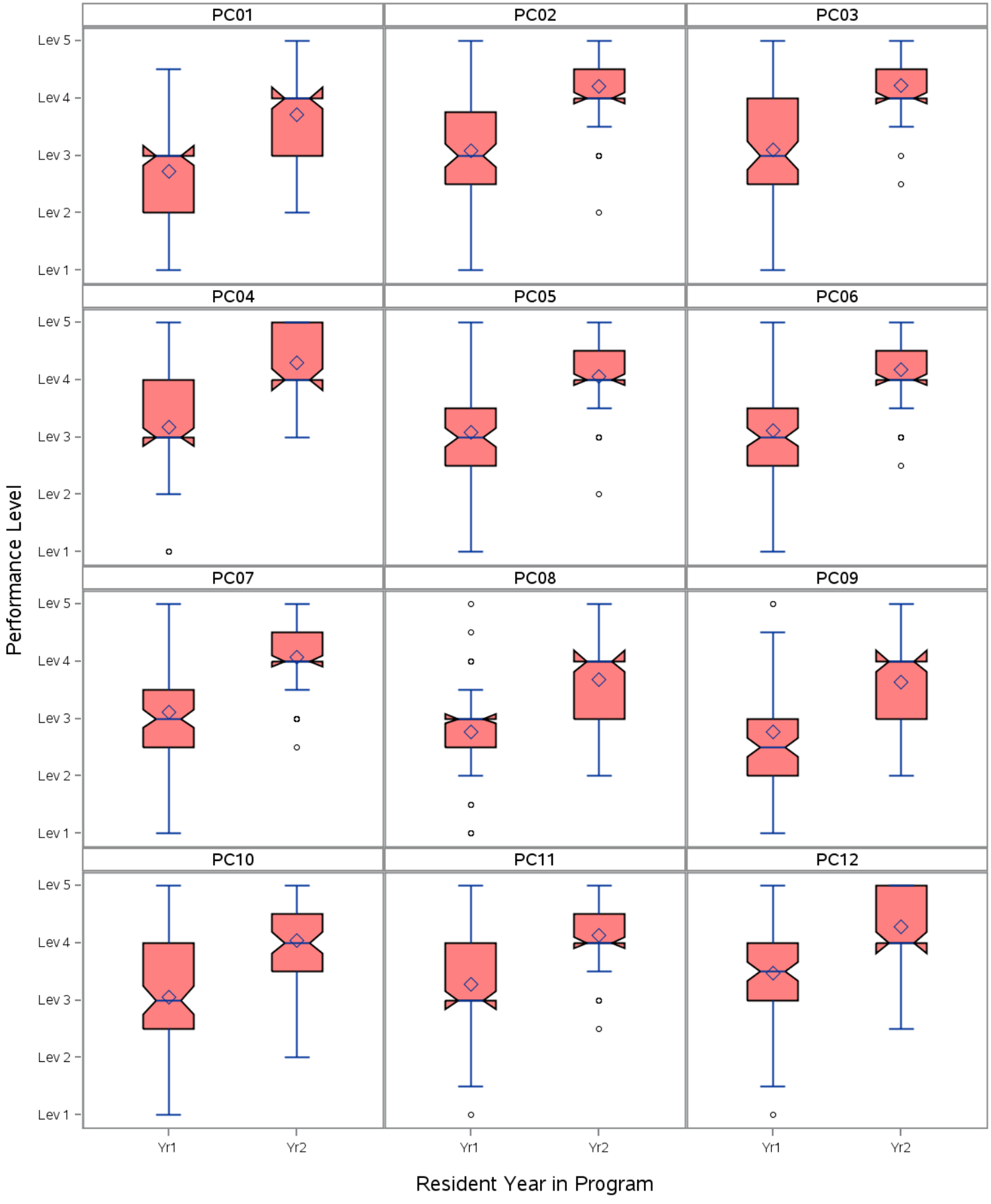


Table 26 - PM_Public Health and General Preventive Medicine (June 2016)

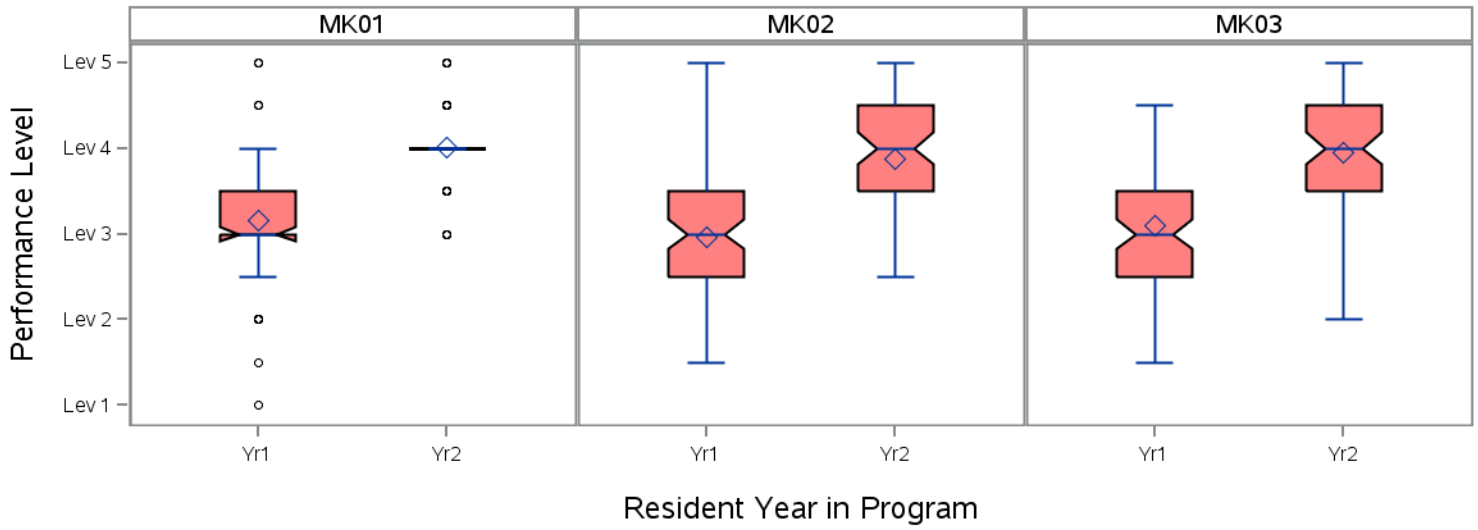


Table 26 - PM_Public Health and General Preventive Medicine (June 2016)

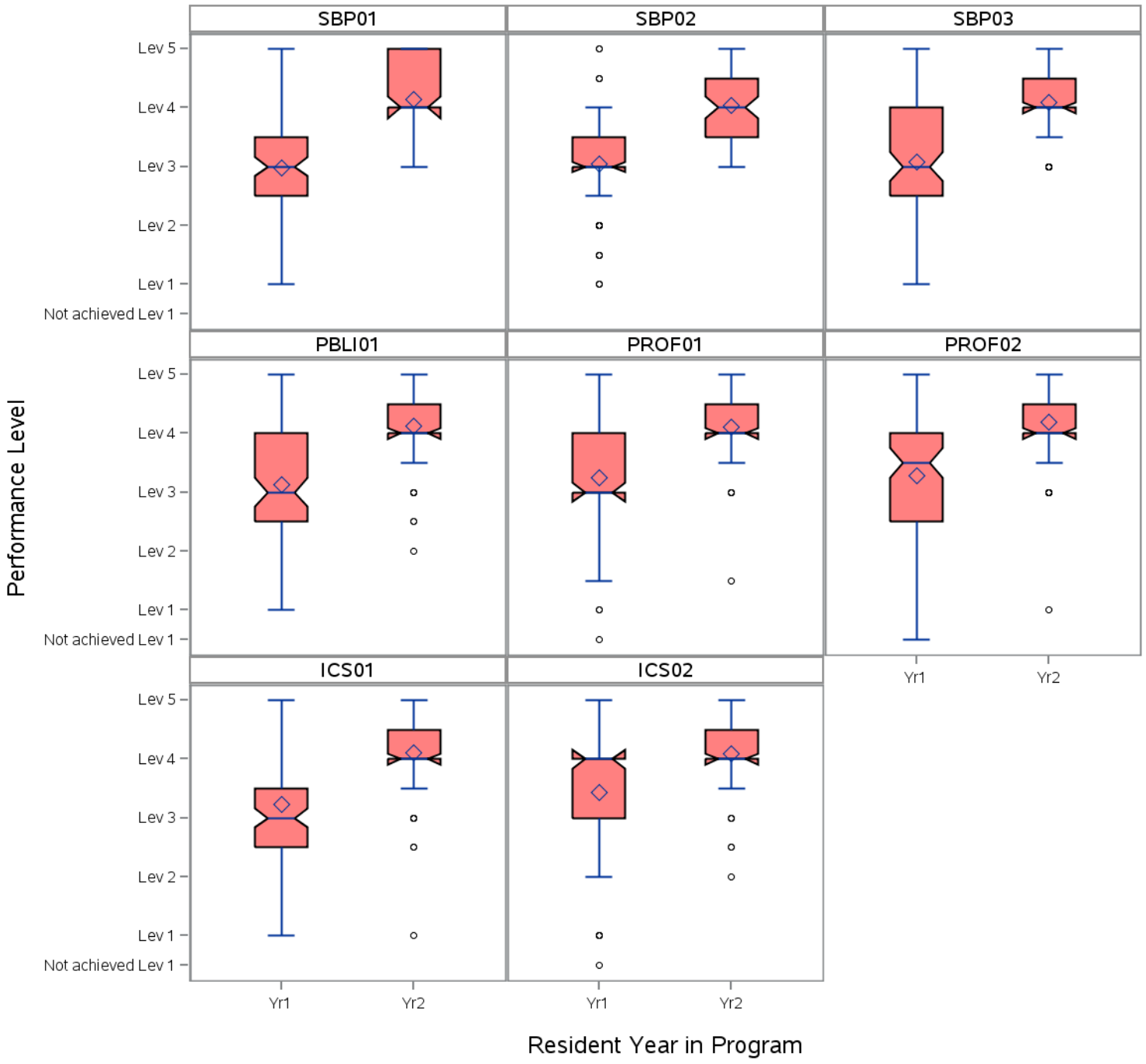


Table 27 - Psychiatry (June 2016)

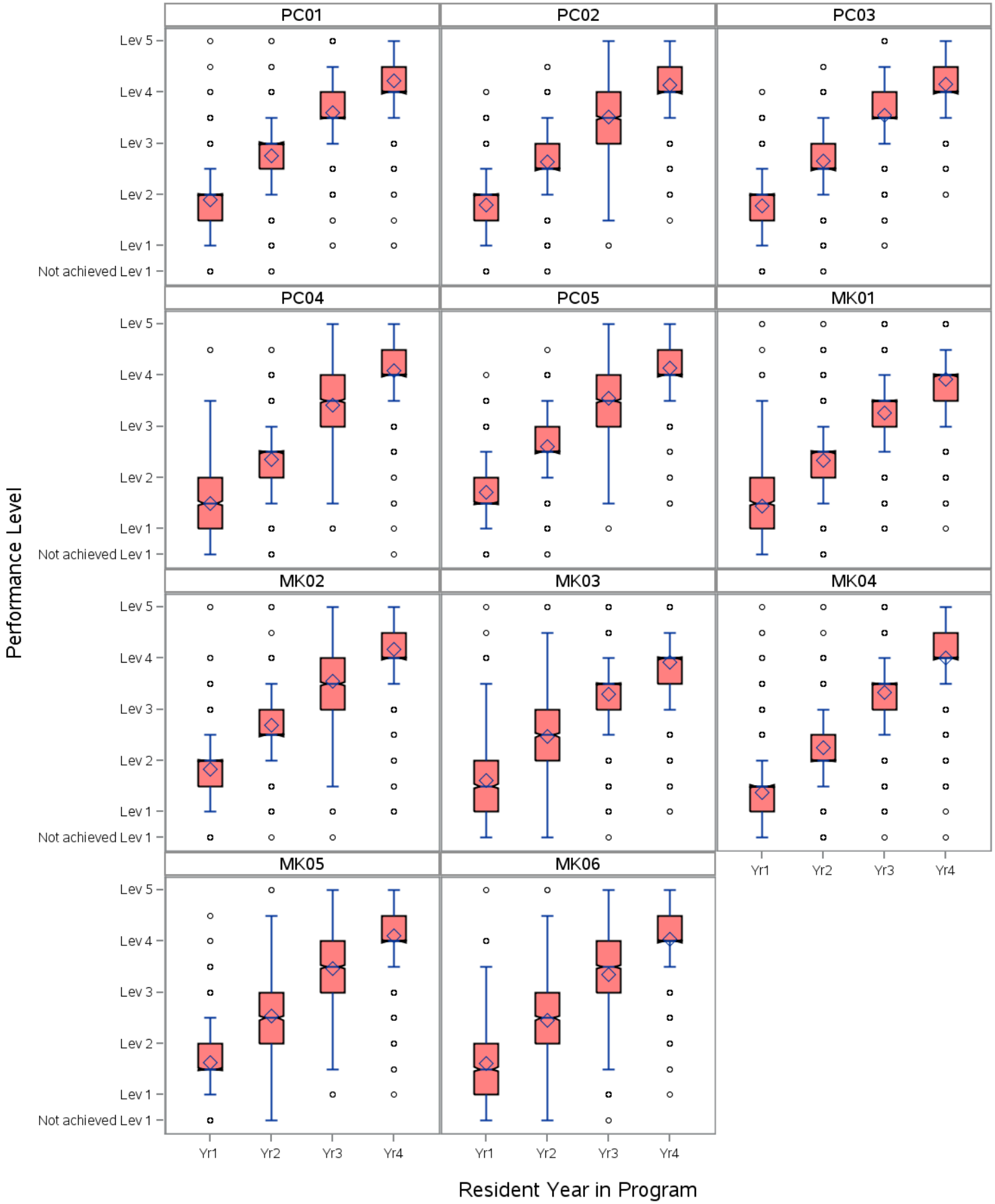


Table 27 - Psychiatry (June 2016)

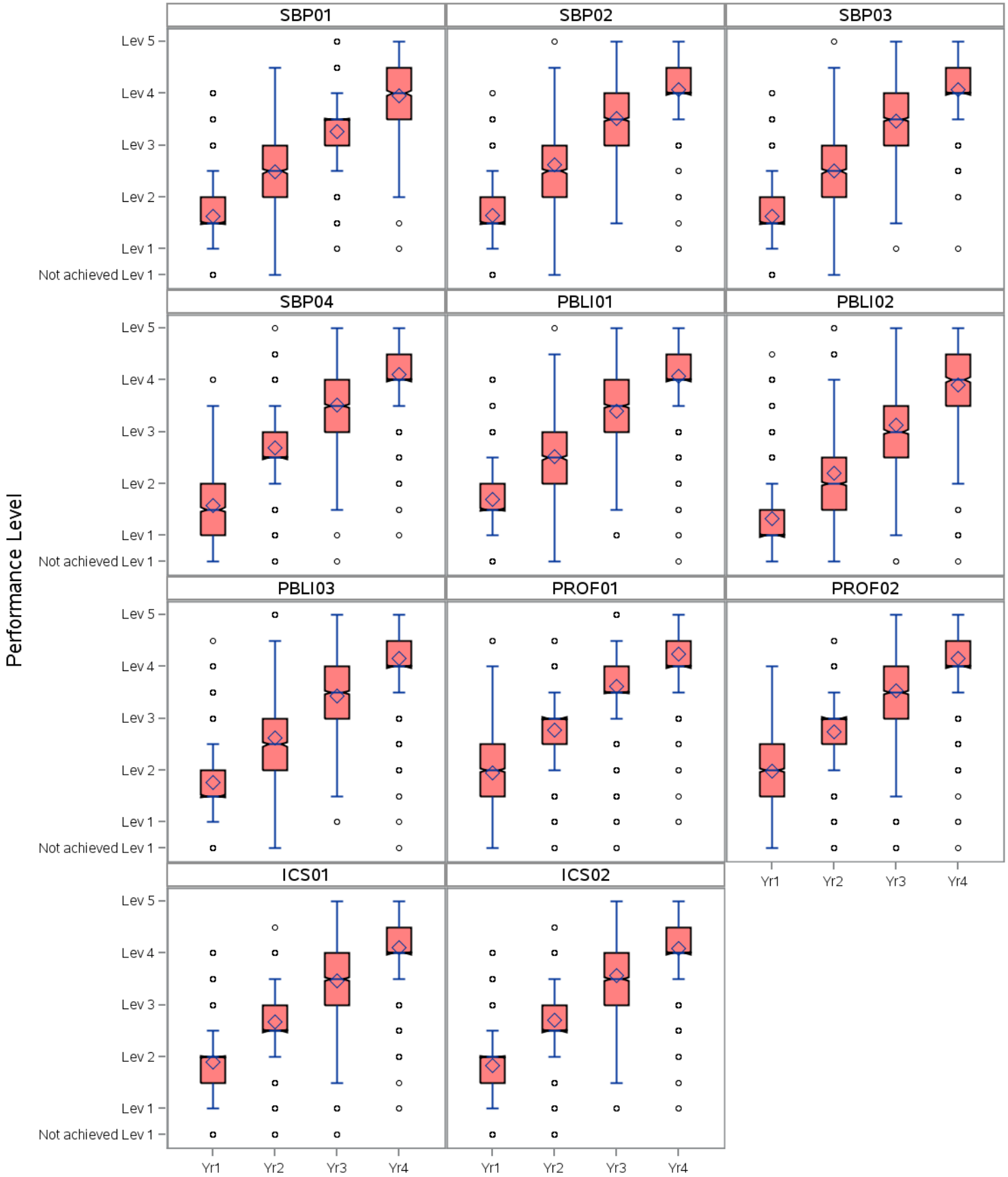


Table 28 - Radiation Oncology (June 2016)

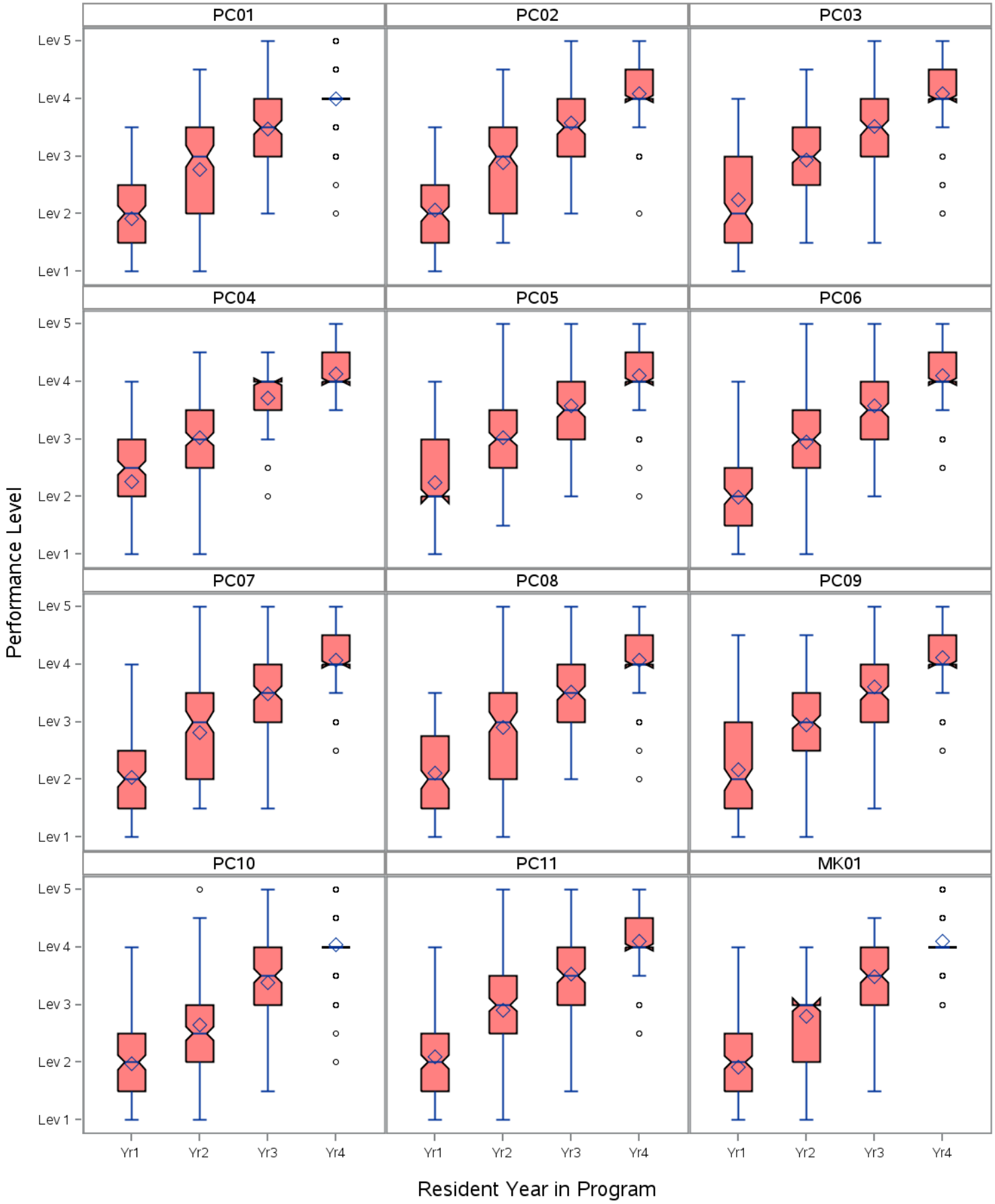


Table 28 - Radiation Oncology (June 2016)

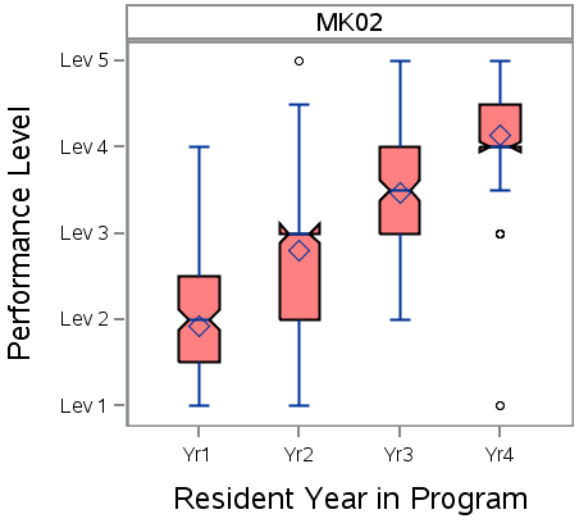


Table 28 - Radiation Oncology (June 2016)

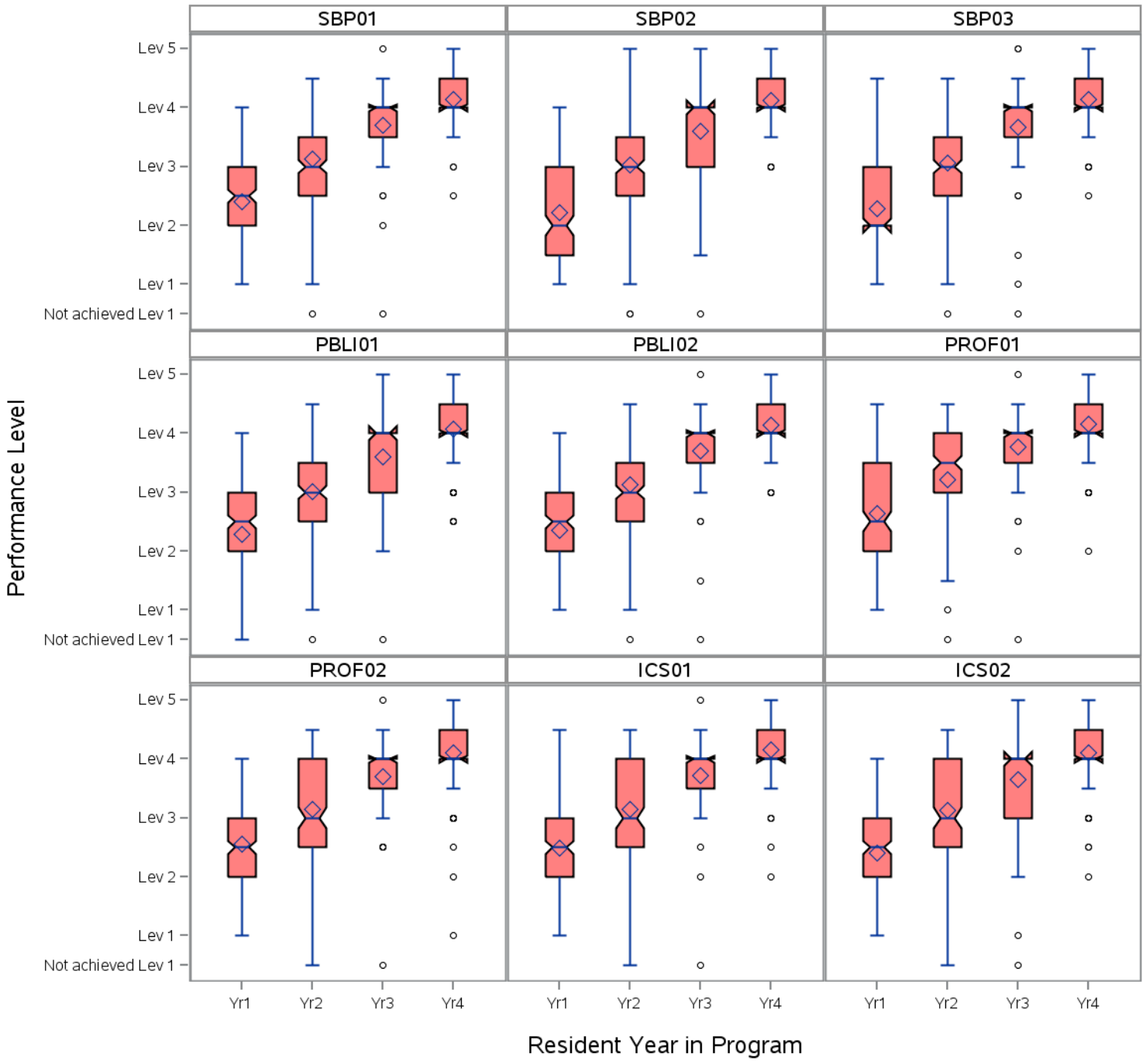


Table 29 - Radiology-Diagnostic (June 2016)

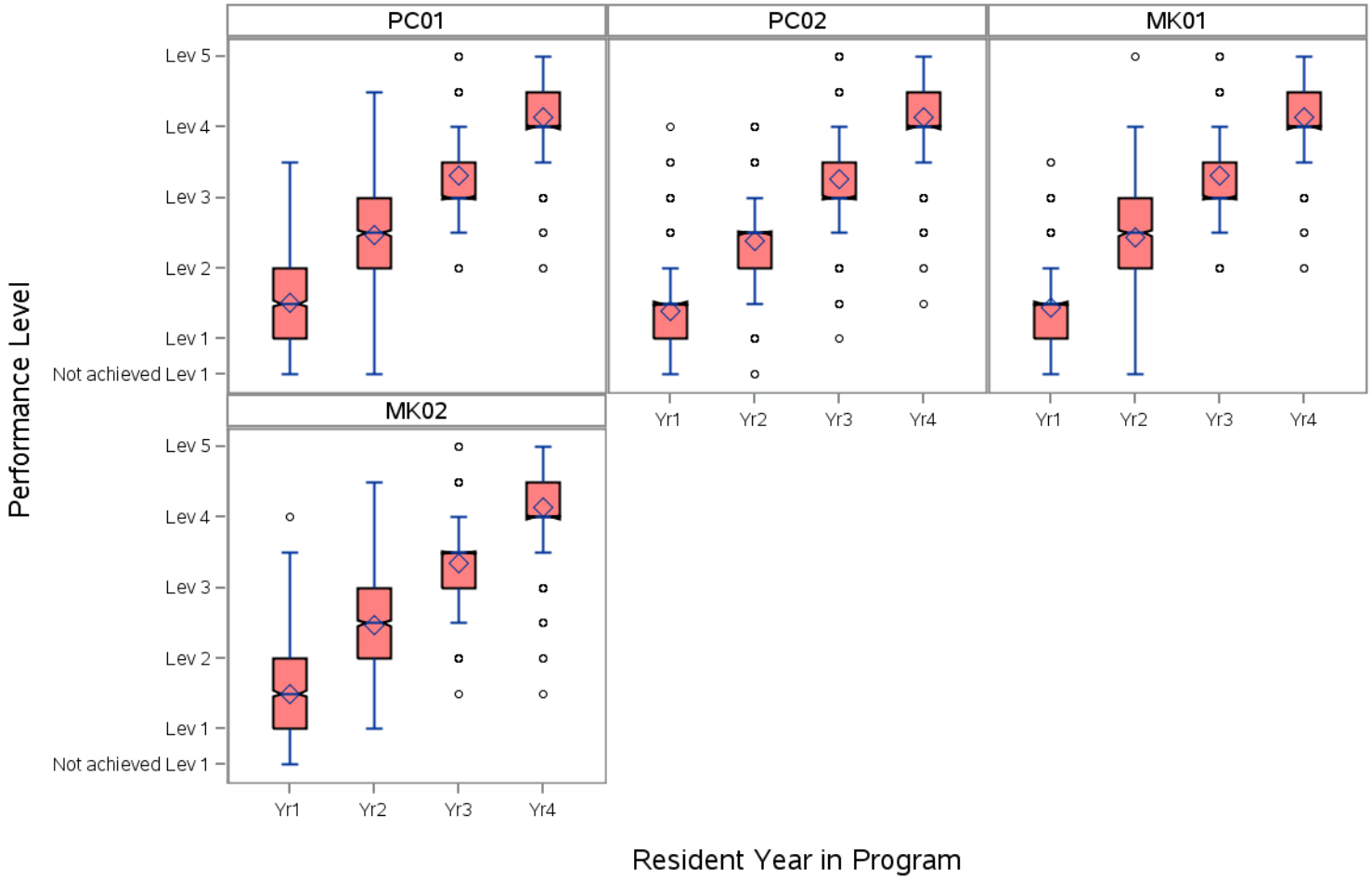


Table 29 - Radiology-Diagnostic (June 2016)

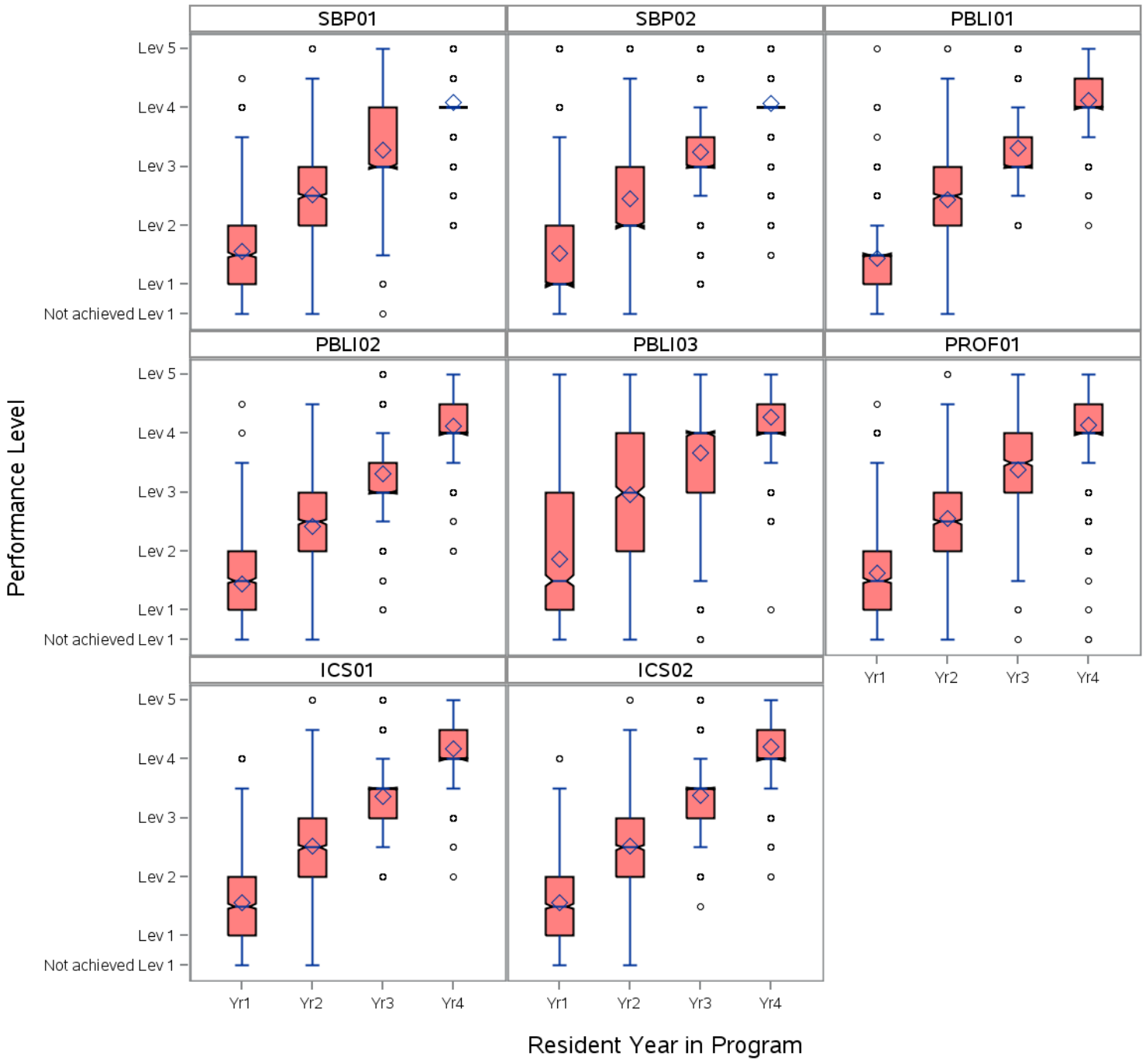


Table 30 - Surgery (June 2016)

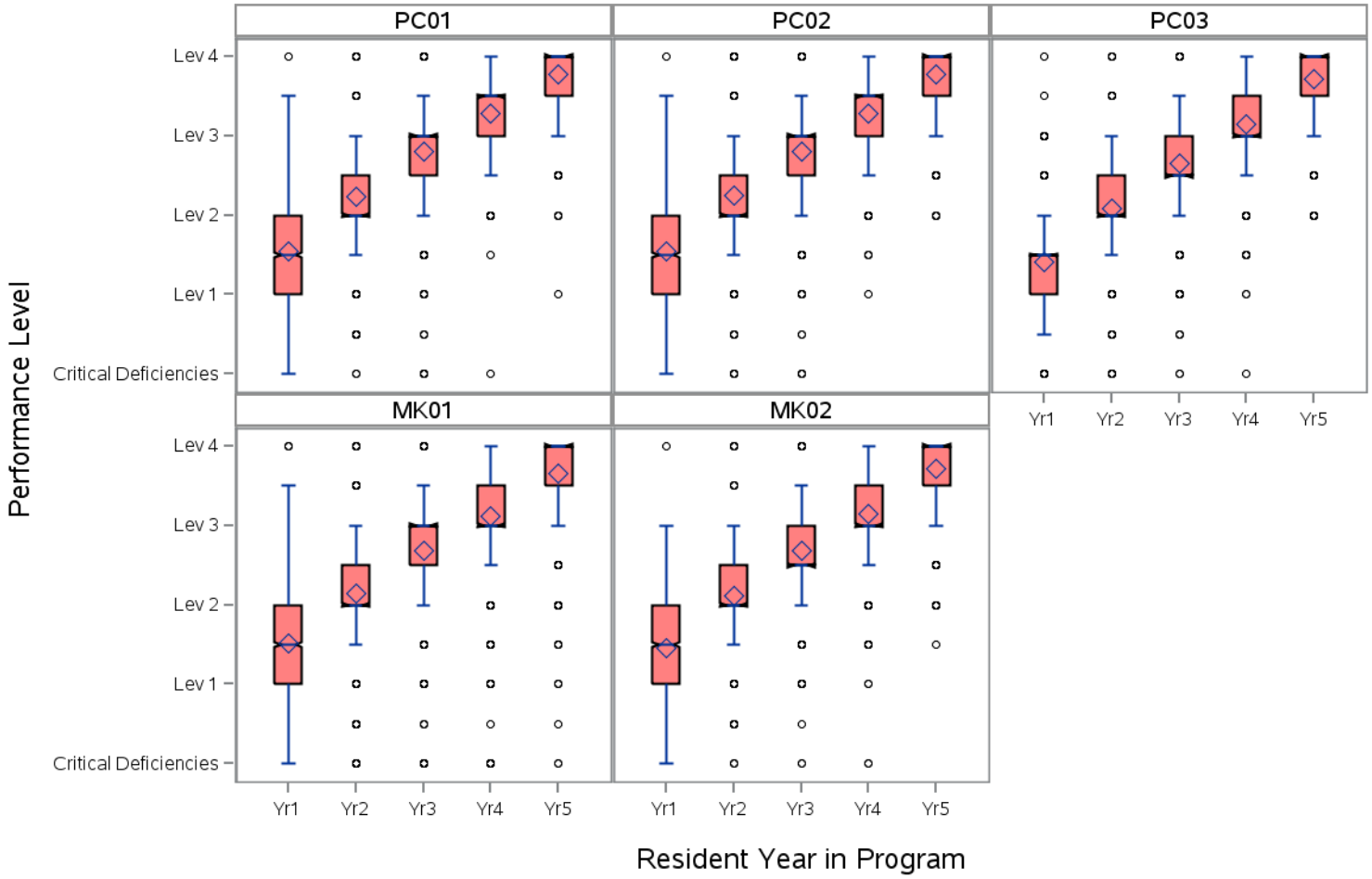


Table 30 - Surgery (June 2016)

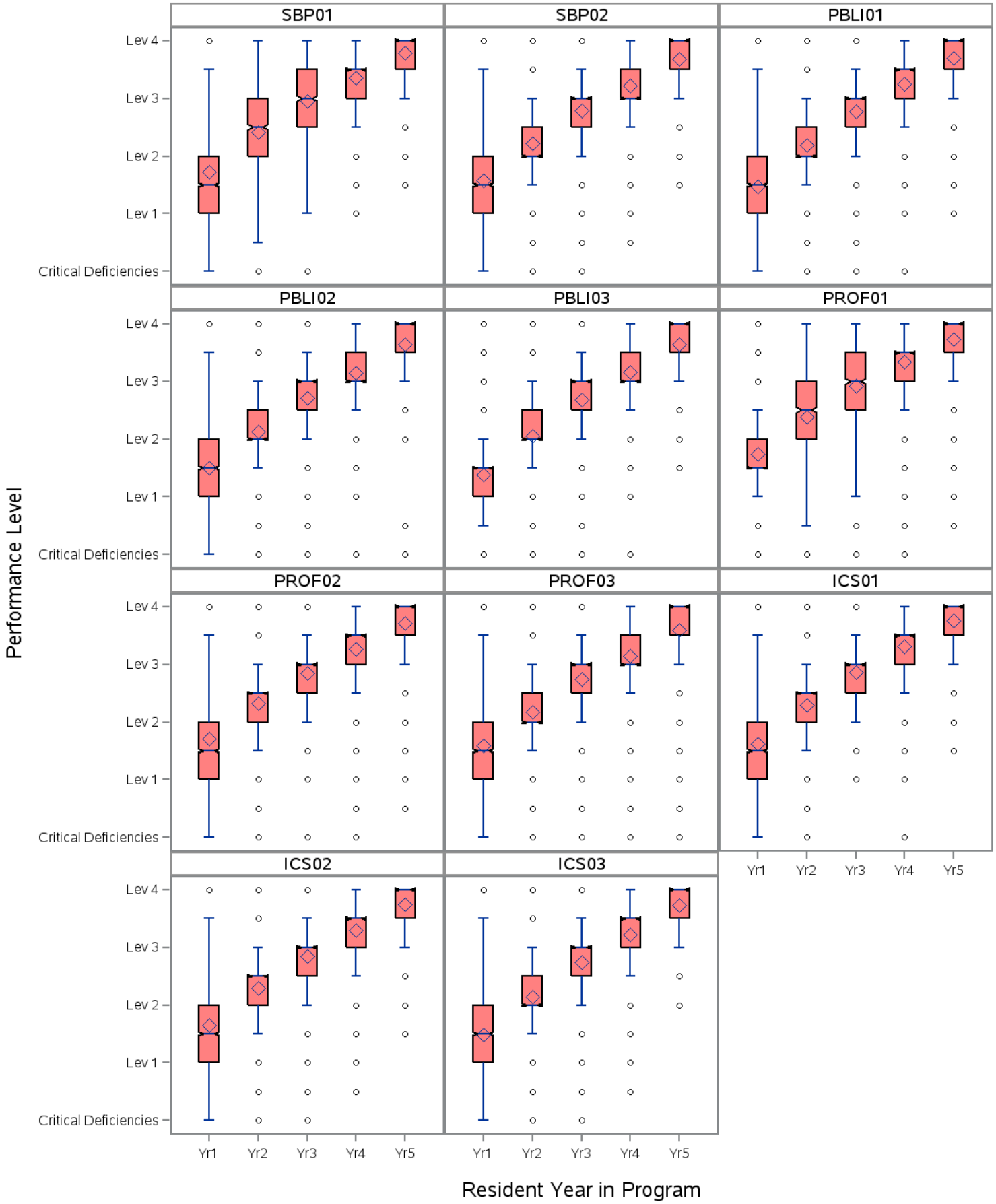


Table 31 - Thoracic Surgery - Integrated (June 2016)

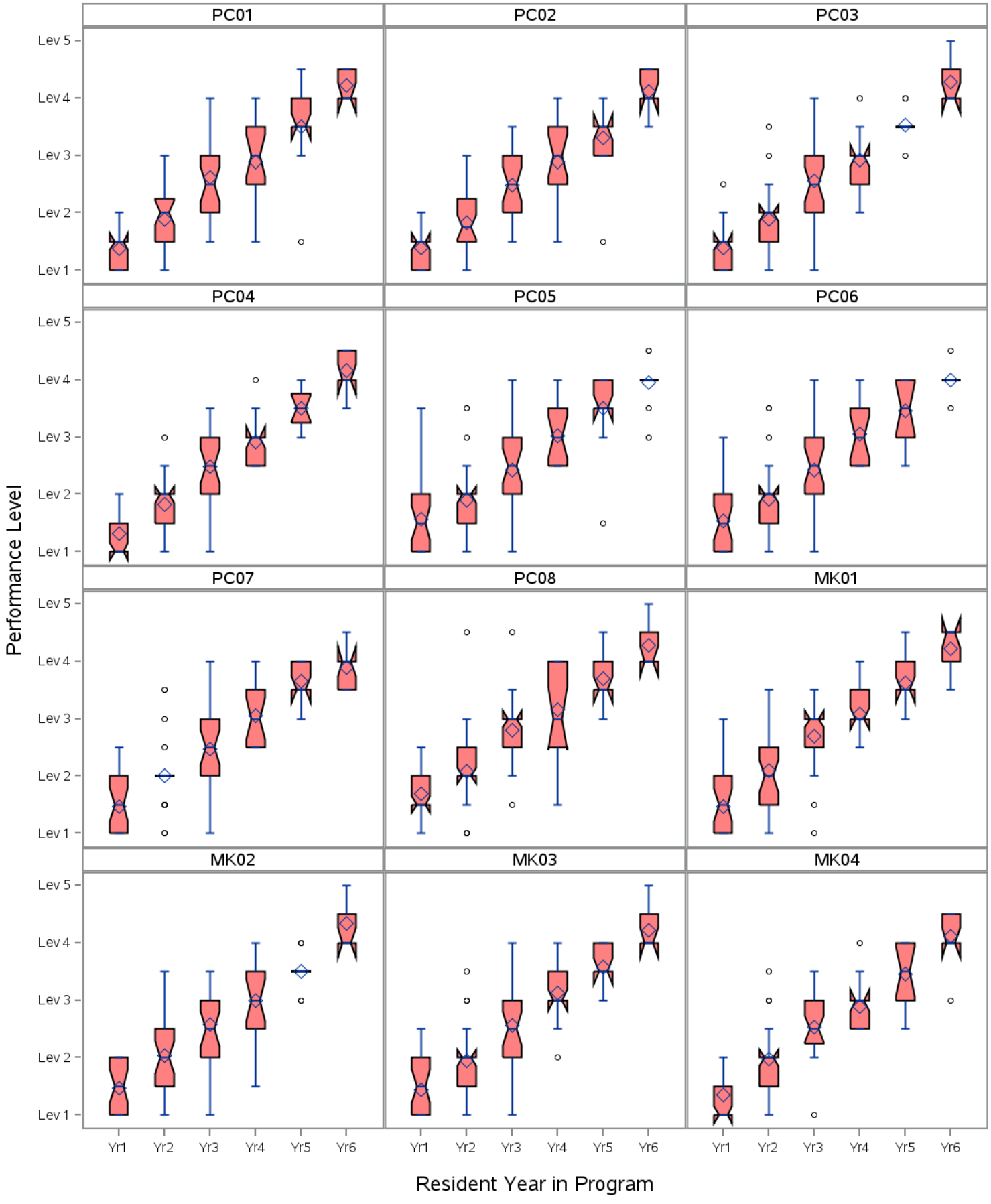


Table 31 - Thoracic Surgery - Integrated (June 2016)

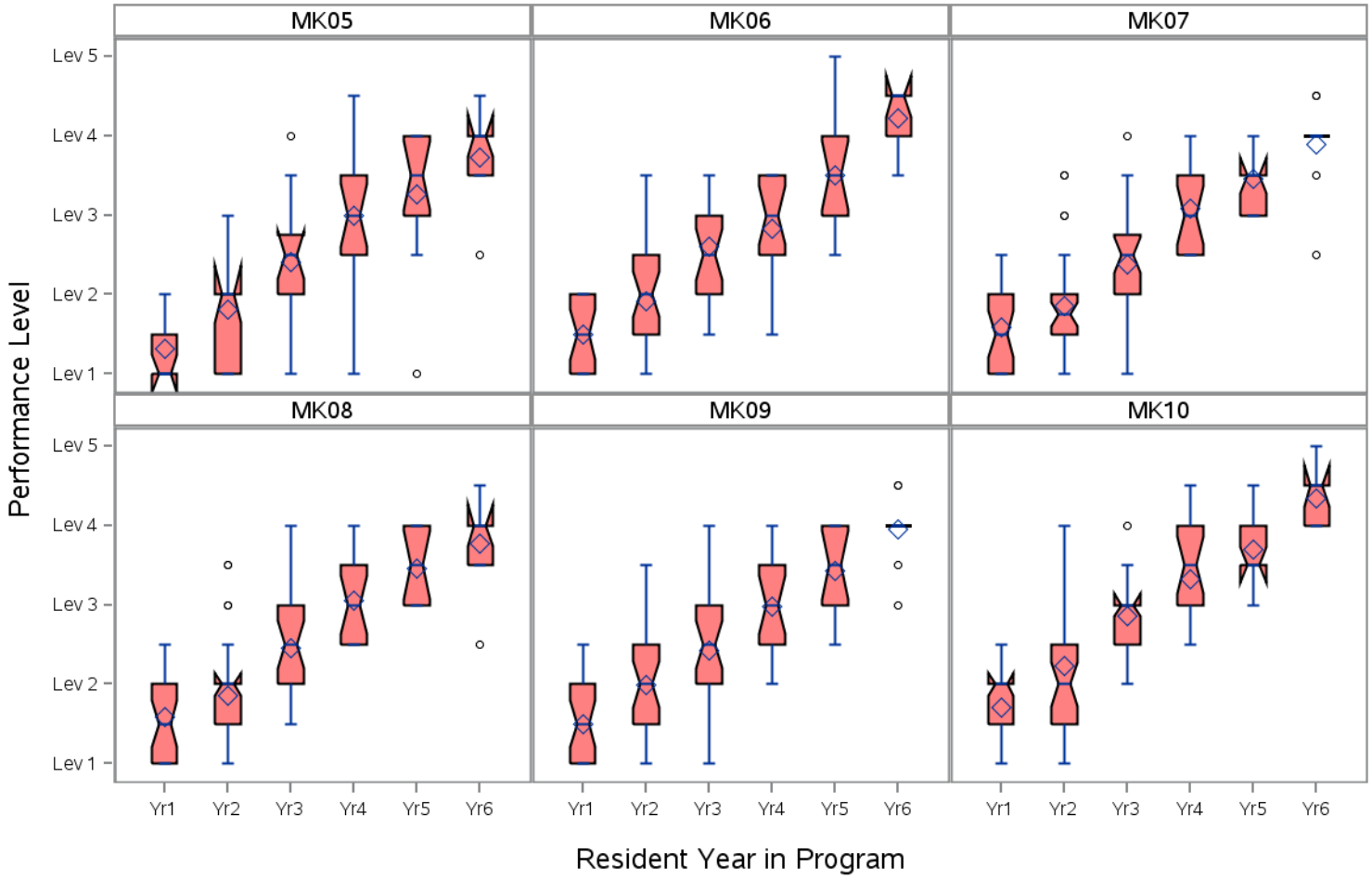


Table 31 - Thoracic Surgery - Integrated (June 2016)

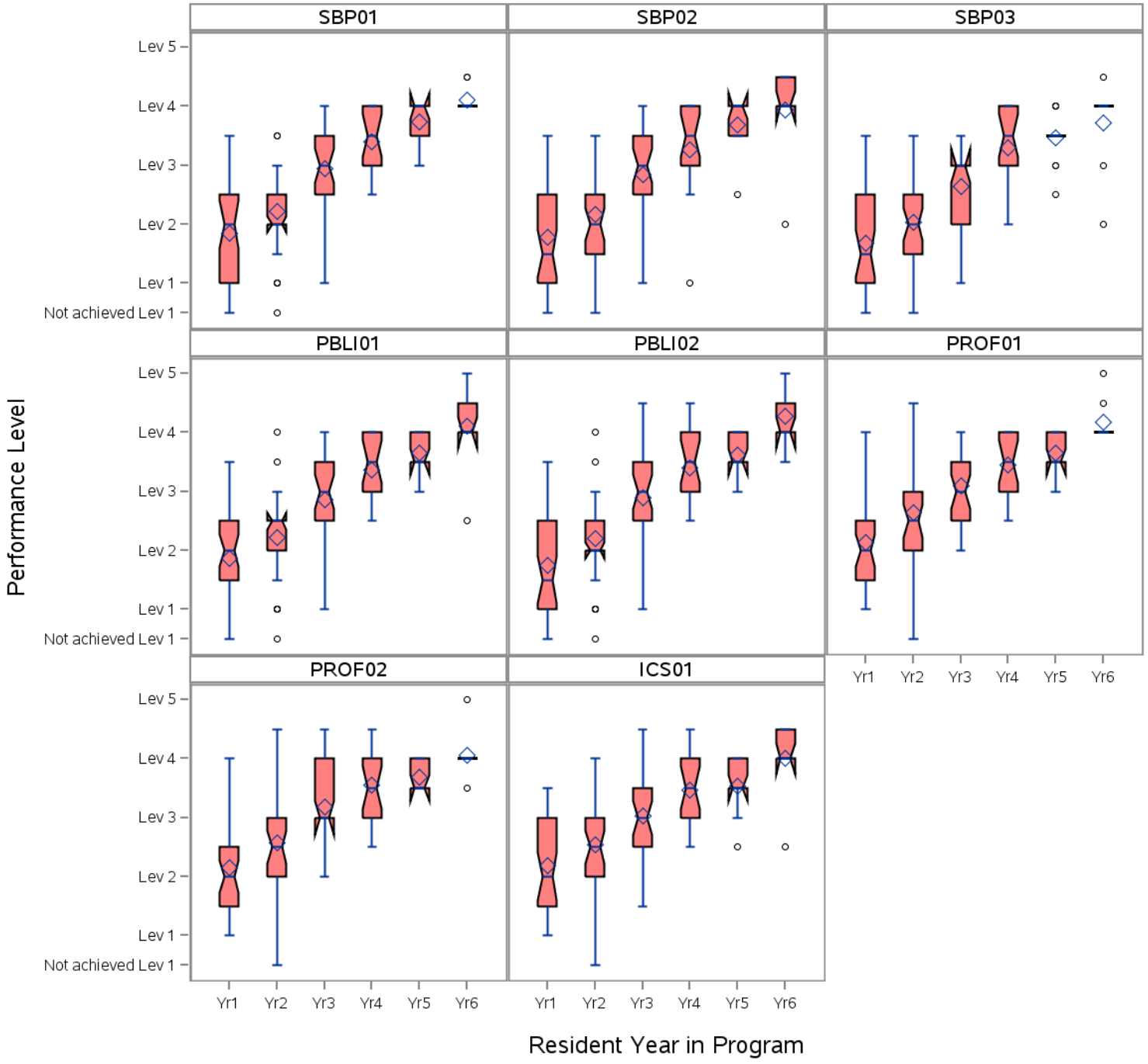


Table 32 - Transitional Year (June 2016)

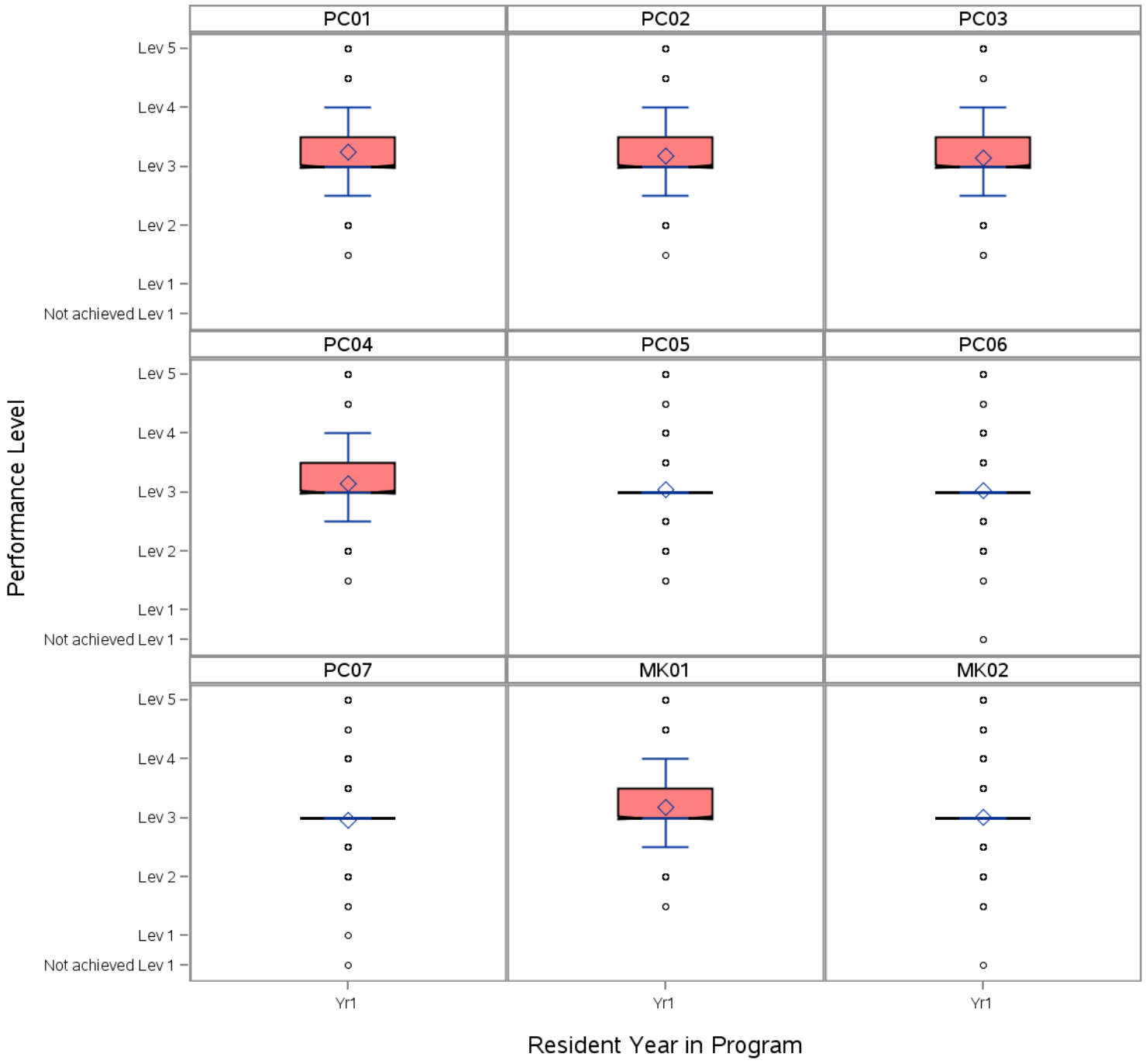


Table 32 - Transitional Year (June 2016)

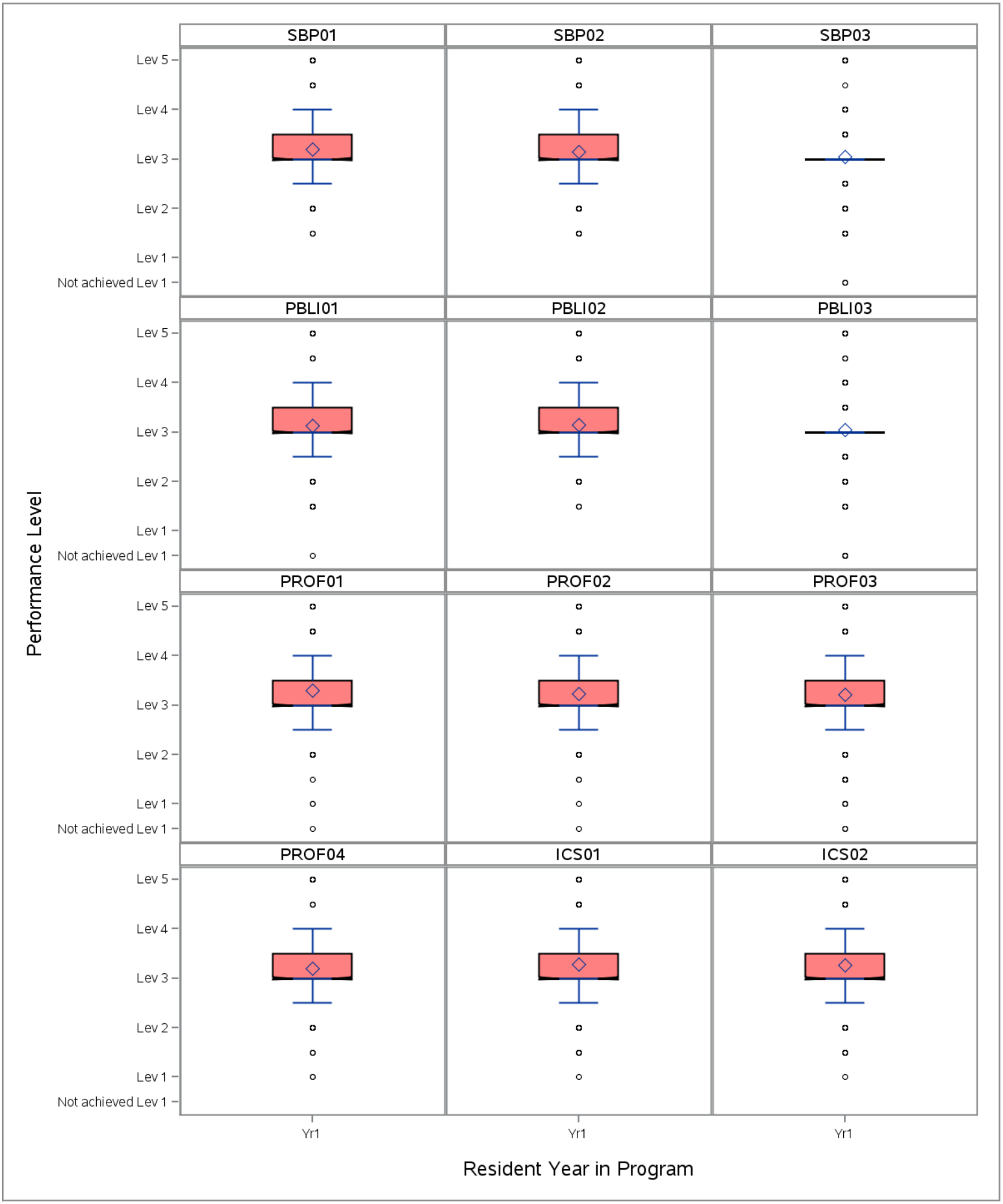


Table 32 - Transitional Year (June 2016)

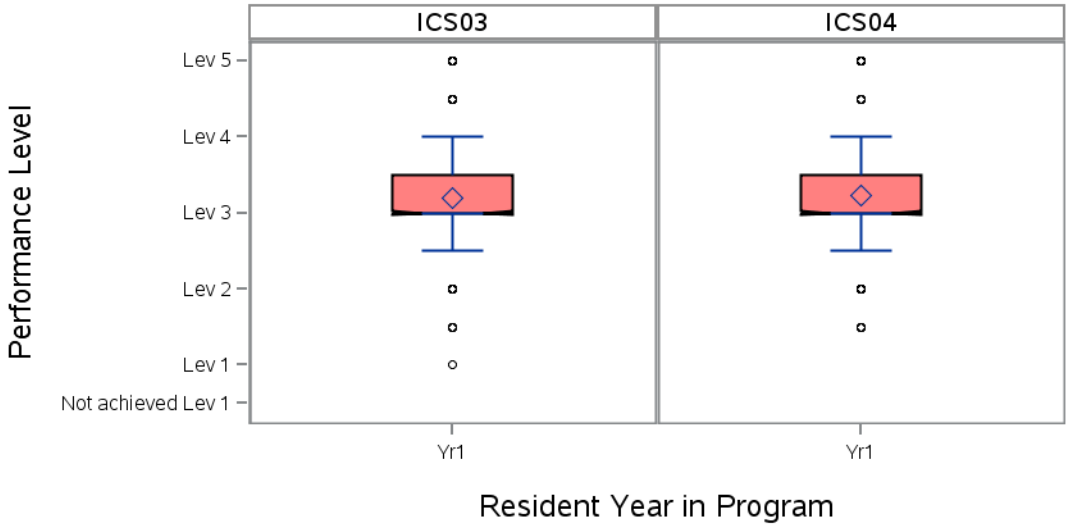


Table 33 - Urology (June 2016)

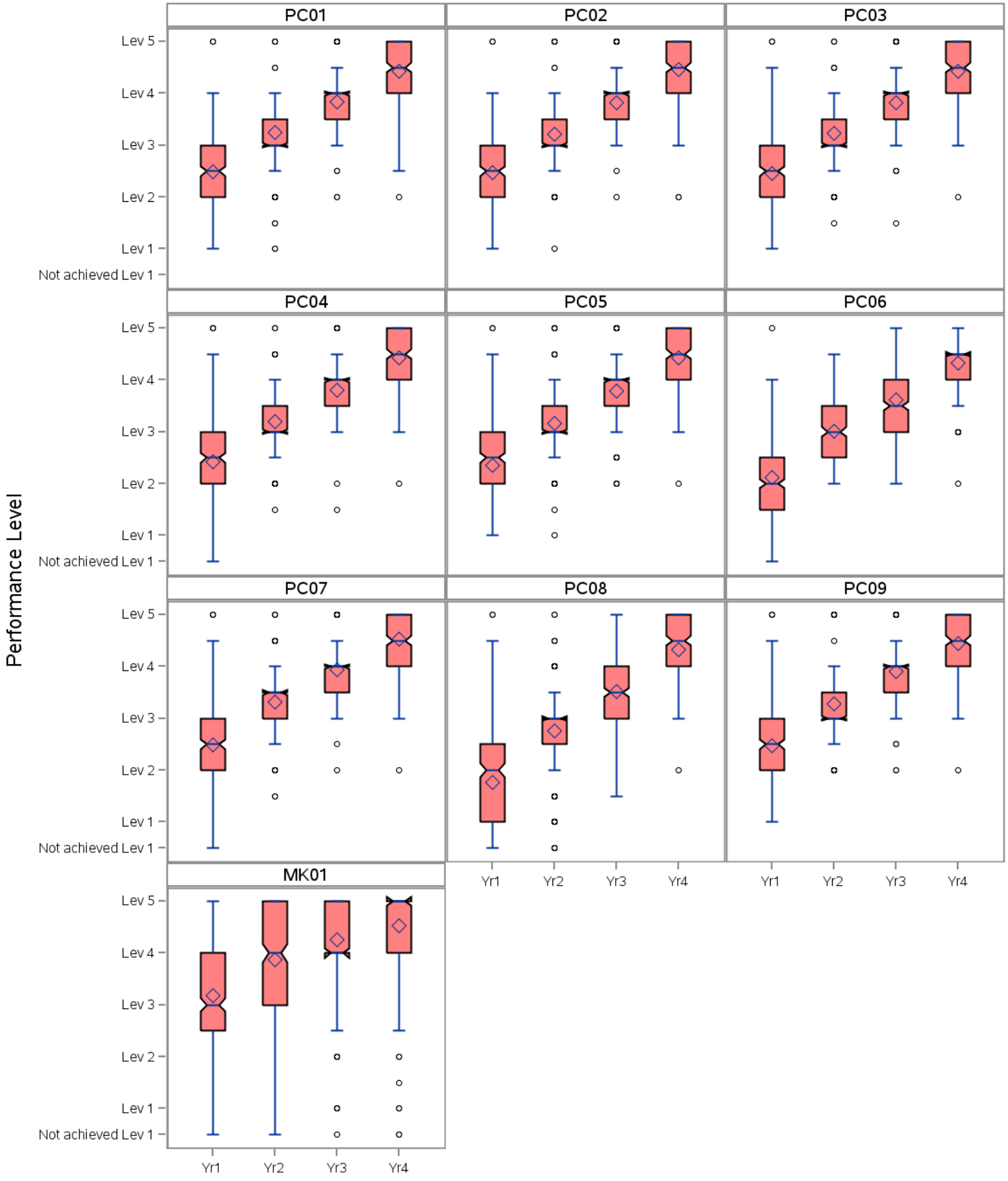


Table 33 - Urology (June 2016)

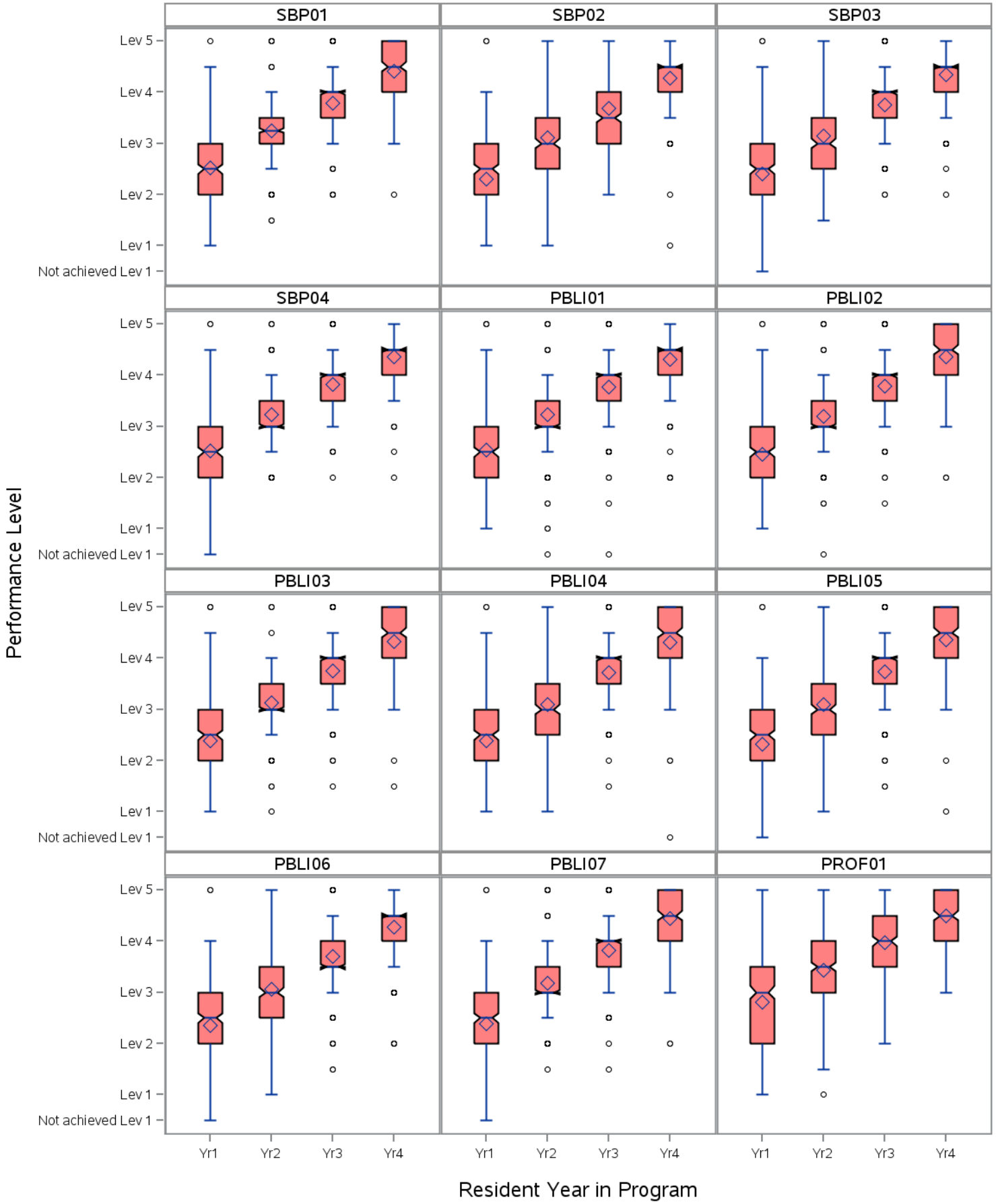


Table 33 - Urology (June 2016)

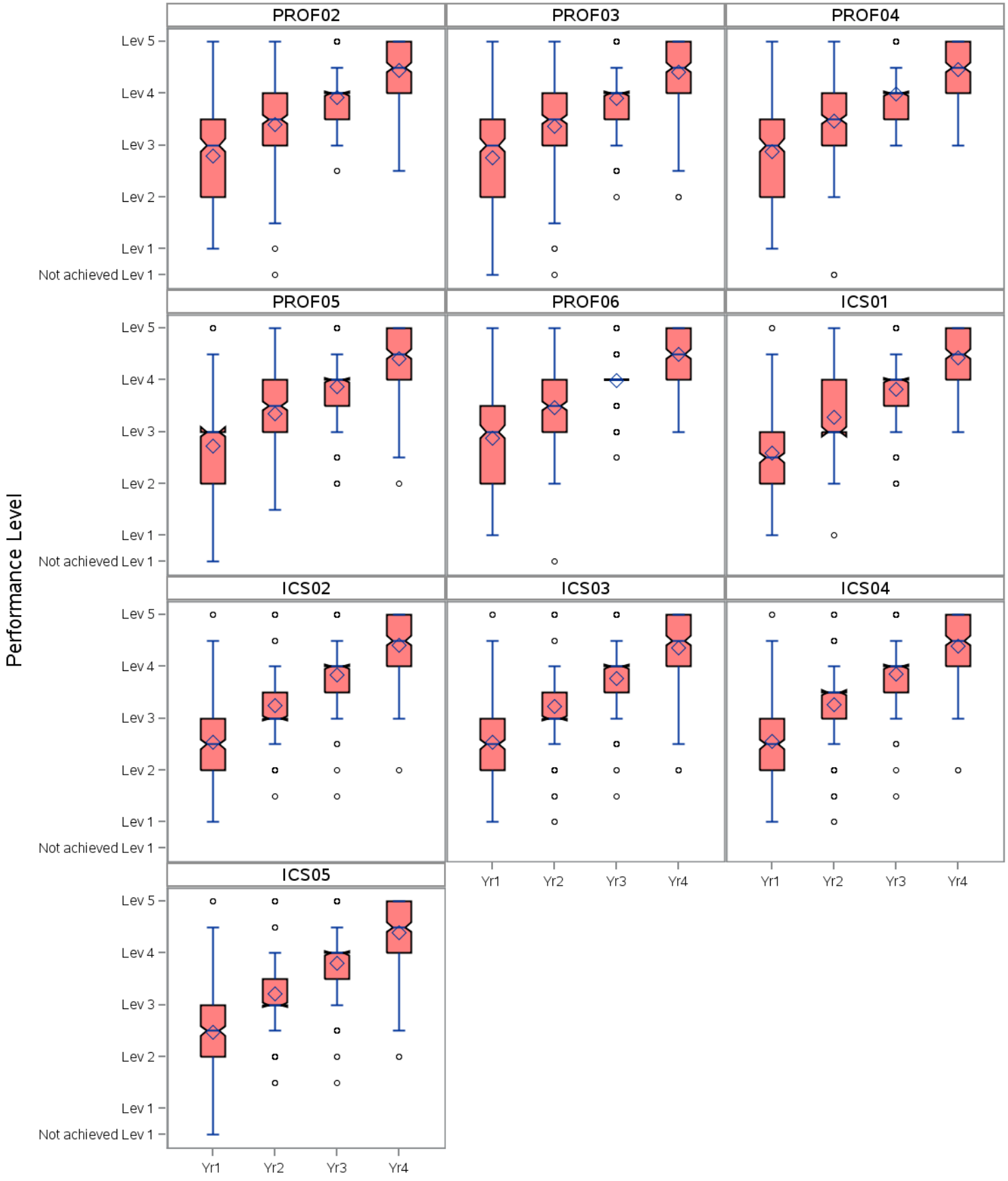


Table 34 - Vascular Surgery - Integrated (June 2016)

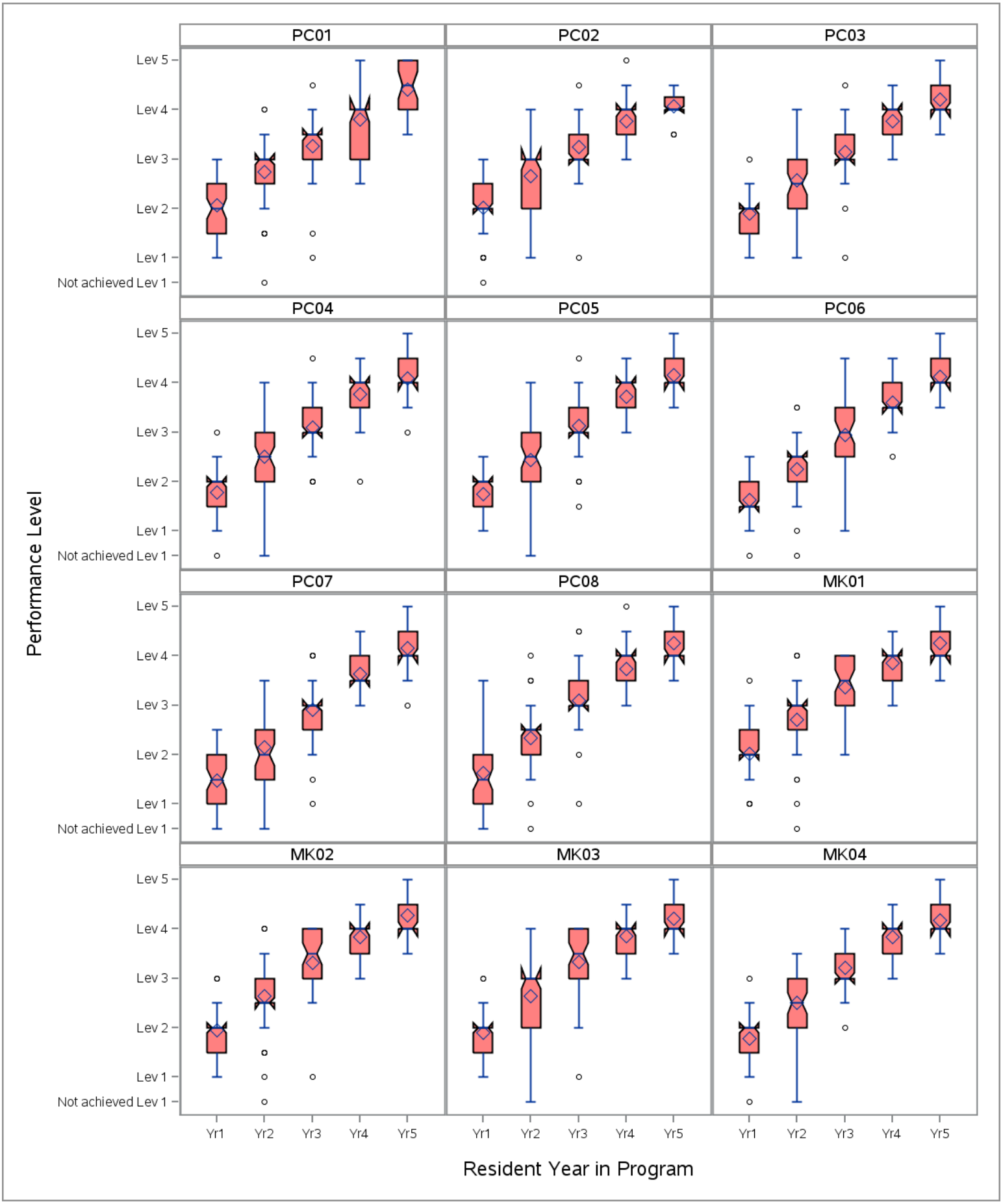


Table 34 - Vascular Surgery - Integrated (June 2016)

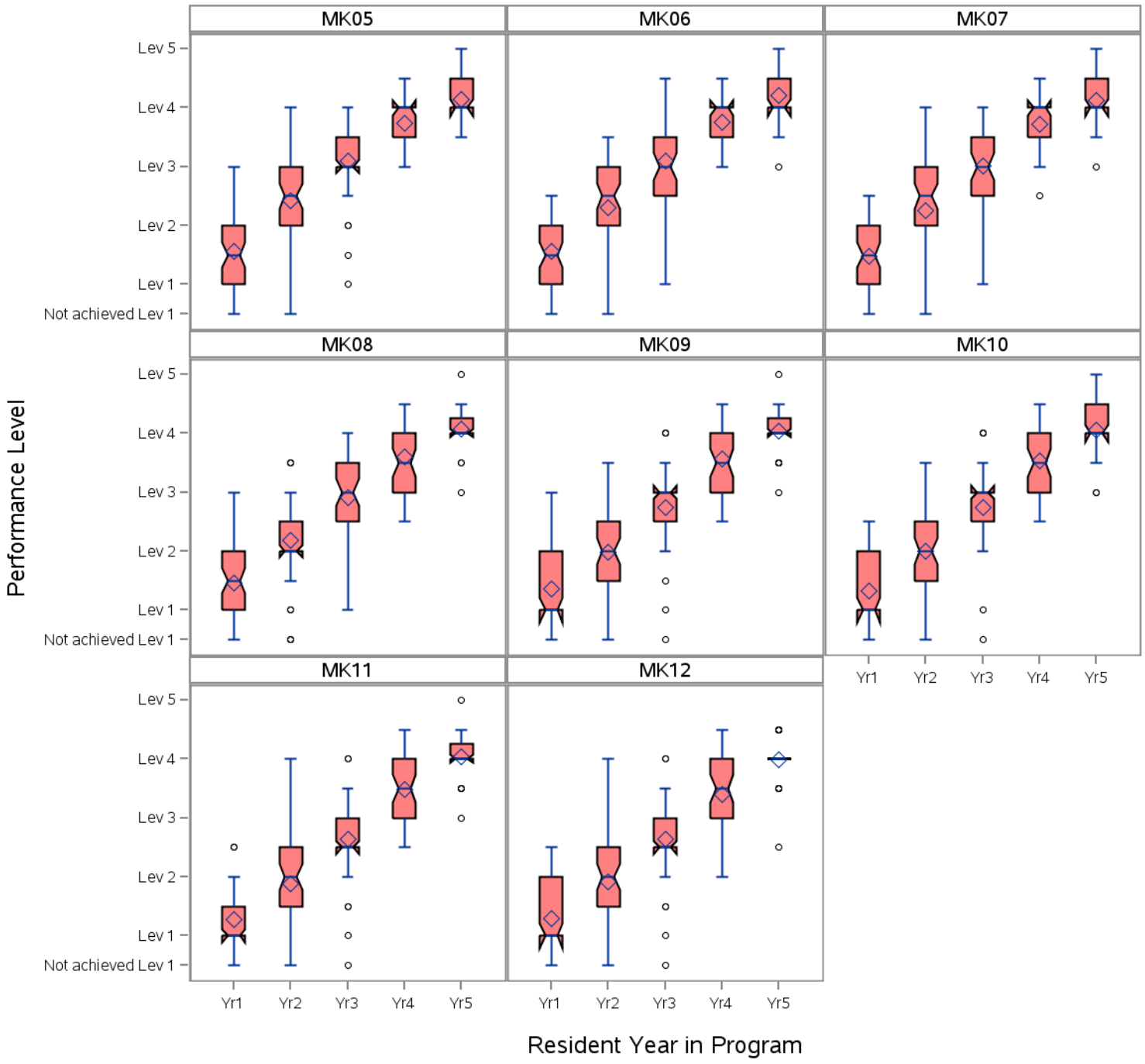
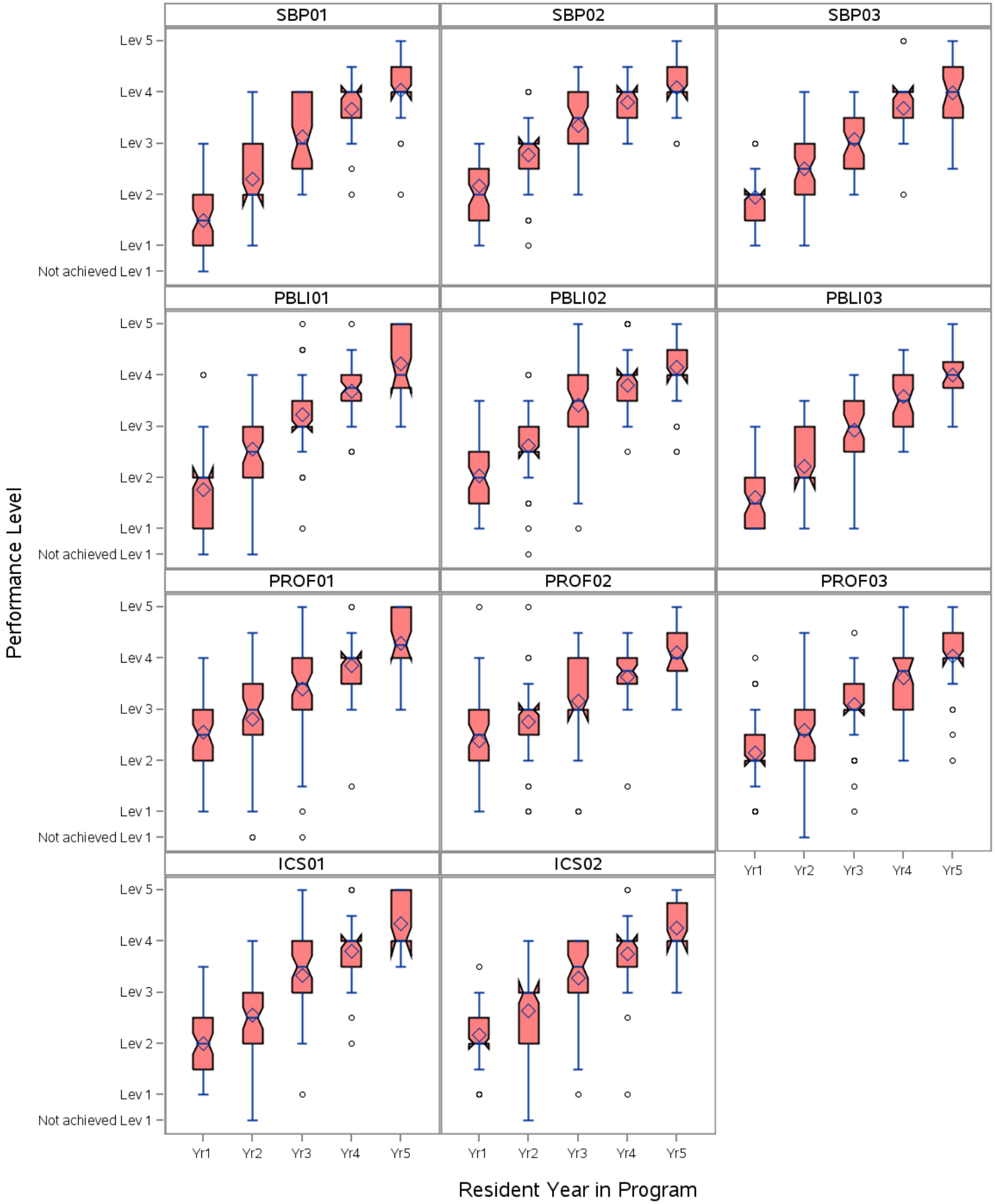


Table 34 - Vascular Surgery - Integrated (June 2016)



Appendix A

Select Milestones Bibliography

National Studies (Full Milestone Set)

1. Beeson MS, Holmboe ES, Korte RC, Nasca TJ, Brigham T, Russ CM, Whitley CT, Reisdorff EJ. Initial Validity Analysis of the Emergency Medicine **Milestones**. Acad Emerg Med. 2015 Jul;22(7):838-44.

OBJECTIVES:
The Accreditation Council for Graduate Medical Education (ACGME) Milestones describe behavioral markers for the progressive acquisition of competencies during residency. As a key component of the Next Accreditation System, all residents are evaluated for the acquisition of specialty-specific Milestones. The objective was to determine the validity and reliability of the emergency medicine (EM) Milestones.
METHODS:
The ACGME and the American Board of Emergency Medicine performed this single-event observational study. The data included the initial EM Milestones performance ratings of all categorical EM residents submitted to the ACGME from October 31, 2013, to January 6, 2014. Mean performance ratings were determined for all 23 subcompetencies for every year of residency training. The internal consistency (reliability) of the Milestones was determined using a standardized Cronbach's alpha coefficient. Exploratory factor analysis was conducted to determine how the subcompetencies were interrelated.
RESULTS:
EM Milestone performance ratings were obtained on 100% of EM residents (n = 5,805) from 162 residency programs. The mean performance ratings of the aggregate and individual subcompetency scores showed discrimination between residency years, and the factor structure further supported the validity of the EM Milestones. The reliability was α = 0.96 within each year of training.
CONCLUSIONS:
The EM Milestones demonstrated validity and reliability as an assessment instrument for competency acquisition. EM residents can be assured that this evaluation process has demonstrated validity and reliability; faculty can be confident that the Milestones are psychometrically sound; and stakeholders can know that the Milestones are a nationally standardized, objective measure of specialty-specific competency acquisition.

2. Korte RC, Beeson MS, Russ CM, Carter WA; Emergency Medicine **Milestones** Working Group, Reisdorff EJ. The emergency medicine **milestones**: a validation study. Acad Emerg Med. 2013 Jul;20(7):730-5.

OBJECTIVES:
The Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties sought to define milestones for skill and knowledge acquisition during residency training. Milestones are significant objective observable events. The milestones are listed within a structure that is derived from the ACGME general competencies. Major groups of milestones are called "subcompetencies." The original 24 subcompetencies containing 255 milestones for emergency medicine (EM) were developed through a multiorganizational group representing most EM stakeholder groups. To assure that the milestones reflected EM resident

progress throughout training, the EM Milestones Working Group (EM MWG) sought to validate the individual milestones.

METHODS:
A computer-based survey was sent to all EM residency programs. The survey period began on April 30, 2012, and concluded on May 15, 2012. Respondents were asked to assign each milestone to a specific level of skill or knowledge acquisition. These levels ranged from a beginning resident to an accomplished clinician. There were two different forms that divided the milestones into two groups of 12 subcompetencies each. Surveys were randomly assigned to programs.
RESULTS:
There were five respondents (the program director and four key faculty) requested from each of the 159 residences. There were responses from 96 programs (60.4%). Of the 795 survey recipients, 28 were excluded due to prior exposure to the EM milestones. Of the remaining 767 potential respondents, 281 completed the survey (36.6%) within a 16-day period. Based on the survey results, the working group adjusted the milestones in the following ways: one entire subcompetency (teaching) was eliminated, six new milestones were created, 34 milestones were eliminated, 26 milestones were reassigned to a lower level score, and 20 were reassigned to a higher level. Nineteen milestones were edited to provide greater clarity. The final result was 227 discrete milestones among 23 subcompetencies.
CONCLUSIONS:
The EM milestones were validated through a milestone assignment process using a computer-based survey completed by program directors and key faculty. Milestones were revised in accordance with the results to better align assignment within each performance level.

3. Hauer KE, Clauser J, Lipner RS, Holmboe ES, Caverzagie K, Hamstra SJ, Hood S, Iobst W, Warm E, McDonald FS. The Internal Medicine Reporting **Milestones**: Cross-sectional Description of Initial Implementation in U.S. **Residency** Programs. Ann Intern Med. 2016 May 10. doi: 10.7326/M15-2411. [Epub ahead of print]

Background:
High-quality assessment of resident performance is needed to guide individual residents' development and ensure their preparedness to provide patient care. To facilitate this aim, reporting milestones are now required across all internal medicine (IM) residency programs.
Objective:
To describe initial milestone ratings for the population of IM residents by IM residency programs.
Design:
Cross-sectional study.
Setting:
IM residency programs.
Participants:
All IM residents whose residency program directors submitted milestone data at the end of the 2013-2014 academic year.
Measurements:
Ratings addressed 6 competencies and 22 subcompetencies. A rating of "not assessable" indicated insufficient information to evaluate the given subcompetency. Descriptive statistics were calculated to describe ratings across competencies and training years.
Results:
Data were available for all 21 774 U.S. IM residents from all 383 programs. Overall, 2889 residents (1621 in postgraduate year 1 [PGY-1], 902 in PGY-2, and 366 in PGY-3) had at least 1 subcompetency rated as not assessable. Summaries of average ratings by competency and

training year showed higher ratings for PGY-3 residents in all competencies. Overall ratings for each of the 6 individual competencies showed that fewer than 1% of third-year residents were rated as "unsatisfactory" or "conditional on improvement." However, when subcompetency milestone ratings were used, 861 residents (12.8%) who successfully completed training had at least 1 competency with all corresponding subcompetencies graded below the threshold of "readiness for unsupervised practice."

Limitation:

Data were derived from a point in time in the first reporting period in which milestones were used.

Conclusion:

The initial milestone-based evaluations of IM residents nationally suggest that documenting developmental progression of competency is possible over training years. Subcompetencies may identify areas in which residents might benefit from additional feedback and experience. Future work is needed to explore how milestones are used to support residents' development and enhance residency curricula.

Multi-institutional Studies

1. Li ST, Tancredi DJ, Schwartz A, Guillot AP, Burke AE, Trimm RF, Guralnick S, Mahan JD, Gifford KA; Association of Pediatric Program Directors (APPD) Longitudinal Educational Assessment Research Network (LEARN) Validity of Resident Self-Assessment Group. Competent for Unsupervised Practice: Use of Pediatric **Residency** Training **Milestones** to Assess Readiness. Acad Med. 2016 Jul 26. [Epub ahead of print]

PURPOSE:

To describe clinical skills progression during pediatric residency using the distribution of pediatric milestone assessments by subcompetency and year of training and to determine reasonable milestone expectations at time of graduation.

METHOD:

Multi-institutional cohort study of the milestones reported to the Accreditation Council for Graduate Medical Education for all 21 pediatric subcompetencies. Most subcompetencies were measured using five milestone levels (1 = novice, 2 = advanced beginner, 3 = competent, 4 = proficient, 5 = master); 3 subcompetencies had only four levels defined.

RESULTS:

Milestone assessments for 2,030 pediatric residents in 47 programs during academic year 2013-2014 were obtained. There was significant variation in end-of-year milestone ratings for residents within each level of training, which decreased as training level increased. Most (78.9%; 434/550) graduating third-year pediatric residents received a milestone rating of ≥ 3 in all 21 subcompetencies; fewer (21.1%; 116/550) received a rating of ≥ 4 in all subcompetencies. Across all training levels, professionalism and interpersonal communication skills were rated highest; quality improvement was rated lowest.

CONCLUSIONS:

Trainees entered residency with a wide range of skills. As they advanced, skill variability within a training level decreased. Most graduating pediatric residents were still advancing on the milestone continuum toward proficiency and mastery, and an expectation of milestone ratings ≥ 4 in all categories upon graduation is unrealistic; milestone ratings ≥ 3 upon graduation may be more realistic. Understanding current pediatric residents' and graduates' skills can help to identify key areas that should be specifically targeted during training.

2. Santen SA, Rademacher N, Heron SL, Khandelwal S, Hauff S, Hopson L. How competent are emergency medicine interns for level 1 **milestones**: who is responsible? Acad Emerg Med. 2013 Jul;20(7):736-9

OBJECTIVES:
The Next Accreditation System (NAS) of the Accreditation Council for Graduate Medical Education (ACGME) includes the implementation of developmental milestones for each specialty. The milestones include five progressively advancing skill levels, with Level 1 defining the skill level of a medical student graduate, and Level 5, that of an attending physician. The goal of this study was to query interns on how well they thought their medical school had prepared them to meet the proposed emergency medicine (EM) Level 1 milestones.
METHODS:
In July 2012, an electronic survey was distributed to the interns of 13 EM residency programs, asking interns whether they were taught and assessed on the proposed Level 1 milestones.
RESULTS:
Of possible participants, 113 of 161 interns responded (70% response rate). The interns represented all four regions of the country. The interns responded that the rates of Level 1 milestones they had been taught ranged from 61% for ultrasound to 98% for performance of focused history and physical examination. A substantial number of interns (up to 39%) reported no instruction on milestones such as patient disposition, pain management, and vascular access. Graduating medical students were less commonly assessed than taught the milestones. Skills with technology, including "explain the role of the electronic health record and computerized physician order entry," were assessed for only 39% of interns, and knowledge (USMLE) and history and physical were assessed in nearly all interns. Disposition, ultrasound, multitasking, and wound management were assessed less than half of the time.
CONCLUSIONS:
Many entering EM interns may not have had either teaching or assessment on the knowledge, skills, and behaviors making up the Level 1 milestones expected for graduating medical students. Thus, there is a potential gap in the teaching and assessment of EM interns. Based on these findings, it is unclear who will be responsible (medical schools, EM clerkships, or residency programs) for ensuring that medical students entering residency have achieved Level 1 milestones.

3. Weizberg M, Bond MC, Cassara M, Doty C, Seamon J. Have First-Year Emergency Medicine Residents Achieved Level 1 on Care-Based **Milestones**? J Grad Med Educ. 2015 Dec;7(4):589-94

BACKGROUND:
Residents in Accreditation Council for Graduate Medical Education accredited emergency medicine (EM) residencies were assessed on 23 educational milestones to capture their progression from medical student level (Level 1) to that of an EM attending physician (Level 5). Level 1 was conceptualized to be at the level of an incoming postgraduate year (PGY)-1 resident; however, this has not been confirmed.
OBJECTIVES:
Our primary objective in this study was to assess incoming PGY-1 residents to determine what percentage achieved Level 1 for the 8 emergency department (ED) patient care-based milestones (PC 1-8), as assessed by faculty. Secondary objectives involved assessing what percentage of residents had achieved Level 1 as assessed by themselves, and finally, we calculated the absolute differences between self- and faculty assessments.
METHODS:

Incoming PGY-1 residents at 4 EM residencies were assessed by faculty and themselves during their first month of residency. Performance anchors were adapted from ACGME milestones.
RESULTS:
Forty-one residents from 4 programs were included. The percentage of residents who achieved Level 1 for each subcompetency on faculty assessment ranged from 20% to 73%, and on self-assessment from 34% to 92%. The majority did not achieve Level 1 on faculty assessment of milestones PC-2, PC-3, PC-5a, and PC-6, and on self-assessment of PC-3 and PC-5a. Self-assessment was higher than faculty assessment for PC-2, PC-5b, and PC-6.
CONCLUSIONS:
Less than 75% of PGY-1 residents achieved Level 1 for ED care-based milestones. The majority did not achieve Level 1 on 4 milestones. Self-assessments were higher than faculty assessments for several milestones

4. Schwartz A, Margolis MJ, Multerer S, Haftel HM, Schumacher DJ; APPD LEARN–NBME Pediatrics **Milestones** Assessment Group. A multi-source feedback tool for measuring a subset of Pediatrics **Milestones**. Med Teach. 2016 Mar 30:1-8. [Epub ahead of print]

BACKGROUND:
The Pediatrics Milestones Assessment Pilot employed a new multisource feedback (MSF) instrument to assess nine Pediatrics Milestones among interns and subinterns in the inpatient context.
OBJECTIVE:
To report validity evidence for the MSF tool for informing milestone classification decisions.
METHODS:
We obtained MSF instruments by different raters per learner per rotation. We present evidence for validity based on the unified validity framework.
RESULTS:
One hundred and ninety two interns and 41 subinterns at 18 Pediatrics residency programs received a total of 1084 MSF forms from faculty (40%), senior residents (34%), nurses (22%), and other staff (4%). Variance in ratings was associated primarily with rater (32%) and learner (22%). The milestone factor structure fit data better than simpler structures. In domains except professionalism, ratings by nurses were significantly lower than those by faculty and ratings by other staff were significantly higher. Ratings were higher when the rater observed the learner for longer periods and had a positive global opinion of the learner. Ratings of interns and subinterns did not differ, except for ratings by senior residents. MSF-based scales correlated with summative milestone scores.
CONCLUSION:
We obtain moderately reliable MSF ratings of interns and subinterns in the inpatient context to inform some milestone assignments.

5. Page C, Reid A, Coe CL, Carlough M, Rosenbaum D, Beste J, Fagan B, Steinbacher E, Jones G, Newton WP. Learnings From the Pilot Implementation of Mobile Medical Milestones Application. Journal of Graduate Medical Education 2016; In-Press.

Background Implementation of the educational milestones benefits from mobile technology that facilitates ready assessments in the clinical environment. We developed a point-of-care resident evaluation tool, Mobile Medical Milestones Application (M3App), and piloted it in 8 North Carolina family medicine residency programs.
Objective We sought to examine variations we found in the use of the tool across programs and explored the experiences of program directors, faculty, and residents to better understand the perceived benefits and challenges of implementing the new tool.

Methods Residents and faculty completed presurveys and postsurveys about the tool and the evaluation process in their program. Program directors were interviewed individually. Interviews and open-ended survey responses were analyzed and coded using the constant comparative method, and responses were tabulated under themes.
Results Common perceptions included increased data collection, enhanced efficiency, and increased perceived quality of the information gathered with the M3App. Residents appreciated the timely, high-quality feedback they received. Faculty reported becoming more comfortable with the tool over time, and a more favorable evaluation of the tool was associated with higher utilization. Program directors reported improvements in faculty knowledge of the milestones and resident satisfaction with feedback.
Conclusions Faculty and residents credited the M3App with improving the quality and efficiency of resident feedback. Residents appreciated the frequency, proximity, and specificity of feedback, and faculty reported the app improved their familiarity with the milestones. Implementation challenges included lack of a physician champion and competing demands on faculty time.

6. Bradley KE, Andolsek KM. A pilot study of orthopaedic resident self-assessment using a **milestones**' survey just prior to **milestones** implementation. Int J Med Educ. 2016 Jan 11;7:11-8.

OBJECTIVE:
To pilot test if Orthopaedic Surgery residents could self-assess their performance using newly created milestones, as defined by the Accreditation Council on Graduate Medical Education.
METHODS:
In June 2012, an email was sent to Program Directors and administrative coordinators of the 154 accredited Orthopaedic Surgery Programs, asking them to send their residents a link to an online survey. The survey was adapted from the Orthopaedic Surgery Milestone Project. Completed surveys were aggregated in an anonymous, confidential database. SAS 9.3 was used to perform the analyses.
RESULTS:
Responses from 71 residents were analyzed. First and second year residents indicated through self-assessment that they had substantially achieved Level 1 and Level 2 milestones. Third year residents reported they had substantially achieved 30/41, and fourth year residents, all Level 3 milestones. Fifth year, graduating residents, reported they had substantially achieved 17 Level 4 milestones, and were extremely close on another 15. No milestone was rated at Level 5, the maximum possible. Earlier in training, Patient Care and Medical Knowledge milestones were rated lower than the milestones reflecting the other four competencies of Practice Based Learning and Improvement, Systems Based Practice, Professionalism, and Interpersonal Communication. The gap was closed by the fourth year.
CONCLUSIONS:
Residents were able to successfully self-assess using the 41 Orthopaedic Surgery milestones. Respondents' rate improved proficiency over time. Graduating residents report they have substantially, or close to substantially, achieved all Level 4 milestones. Milestone self-assessment may be a useful tool as one component of a program's overall performance assessment strategy.

Single Institution Studies

1. Goldman RH, Tuomala RE, Bengtson JM, Stagg AR. How Effective are New **Milestones** Assessments at Demonstrating Resident Growth? 1 Year of Data. J Surg Educ. 2016 Jul 6. pii: S1931-7204(16)30078-2. doi: 10.1016/j.jsurg.2016.06.009. [Epub ahead of print]

OBJECTIVE: Assessment tools that accrue data for the Accreditation Council for Graduate Medical Education Milestones must evaluate residents across multiple dimensions, including medical knowledge, procedural skills, teaching, and professionalism. Our objectives were to: (1) develop an assessment tool to evaluate resident performance in accordance with the Milestones and (2) review trends in resident achievements during the inaugural year of Milestone implementation.
DESIGN: A novel venue and postgraduate year (PGY) specific assessment tool was built, tested, and implemented for both operating room and labor and delivery "venues." Resident development of competence and independence was captured over time. To account for variable rotation schedules, the year was divided into thirds and compared using two-tailed Fisher's exact test.
SETTING: Brigham and Women's and Massachusetts General Hospitals, Boston MA.
PARTICIPANTS: Faculty evaluators and obstetrics and gynecology residents.
RESULTS: A total of 822 assessments of 44 residents were completed between 9/2014 and 6/2015. The percentage of labor and delivery tasks completed "independently" increased monotonically across the start of all years: 8.4% for PGY-1, 60.3% for PGY-2, 73.7% for PGY-3, and 87.5% for PGY-4. Assessments of PGY-1 residents demonstrated a significant shift toward "with minimal supervision" and "independent" for the management of normal labor ($p = 0.03$). PGY-3 residents demonstrated an increase in "able to be primary surgeon" in the operating room, from 36% of the time in the first 2/3 of the year, to 62.3% in the last 1/3 ($p < 0.01$).
CONCLUSION: Assessment tools developed to assist with Milestone assignments capture the growth of residents over time and demonstrate quantifiable differences in achievements between PGY classes. These tools will allow for targeted teaching opportunities for both individual residents and residency programs

2. Kobraei EM, Bohnen JD, George BC, Mullen JT, Lillemoe KD, Austen WG Jr, Liao EC. Uniting Evidence-Based Evaluation with the **ACGME** Plastic Surgery **Milestones**: A Simple and Reliable Assessment of Resident Operative Performance. Plast Reconstr Surg. 2016; Aug;138(2):349e-57e

BACKGROUND:
Milestones evaluations in plastic surgery reflect a shift toward competency-based training but have created a number of challenges. The authors have incorporated the smartphone application evaluation tool, System for Improving and Measuring Procedural Learning (SIMPL), that was recently developed by a multi-institutional research collaborative. In this pilot study, the authors hypothesize that SIMPL can improve resident evaluation and also collect granular performance data to simplify compliance with the plastic surgery Milestones.
METHODS:
SIMPL was prospectively piloted with a plastic surgery resident and faculty surgeon at Massachusetts General Hospital in this institutional review board-approved study. The study period was a 2-month interval corresponding to the resident's rotation.
RESULTS:
The resident-faculty combination performed 20 cases together. All cases were evaluated with SIMPL. SIMPL evaluations uniformly took under 1 minute to submit. The average time to

completed evaluation from surgery completion was 5 hours (<0.5 hour to 12 hours). Concrete, objective, and specific data about resident performance were collected for every case and presented in a concise format.
CONCLUSIONS:
SIMPL is an innovative, evidence-based evaluation system that makes performance assessment feasible for every procedure in which a plastic surgery resident participates. SIMPL's competency-based framework can be easily scaled to facilitate data collection and reporting of mandatory Milestones evaluations at the program and national levels. This technology will support a shared vocabulary between residents and faculty to enhance intraoperative education

3. Choe JH, Knight CL, Stiling R, Corning K, Lock K, Steinberg KP. Shortening the Miles to the **Milestones**: Connecting EPA-Based Evaluations to **ACGME** Milestone Reports for Internal Medicine Residency Programs. Acad Med. 2016 Jul;91(7):943-50.

The Next Accreditation System requires internal medicine training programs to provide the Accreditation Council for Graduate Medical Education (ACGME) with semiannual information about each resident's progress in 22 subcompetency domains. Evaluation of resident "trustworthiness" in performing entrustable professional activities (EPAs) may offer a more tangible assessment construct than evaluations based on expectations of usual progression toward competence. However, translating results from EPA-based evaluations into ACGME milestone progress reports has proven to be challenging because the constructs that underlay these two systems differ. The authors describe a process to bridge the gap between rotation-specific EPA-based evaluations and ACGME milestone reporting. Developed at the University of Washington in 2012 and 2013, this method involves mapping EPA-based evaluation responses to "milestone elements," the narrative descriptions within the columns of each of the 22 internal medicine subcompetencies. As faculty members complete EPA-based evaluations, the mapped milestone elements are automatically marked as "confirmed." Programs can maintain a database that tallies the number of times each milestone element is confirmed for a resident; these data can be used to produce graphical displays of resident progress along the internal medicine milestones. Using this count of milestone elements allows programs to bridge the gap between faculty assessments of residents based on rotation-specific observed activities and semiannual ACGME reports based on the internal medicine milestones. Although potentially useful for all programs, this method is especially beneficial to large programs where clinical competency committee members may not have the opportunity for direct observation of all residents.

4. Goldflam K, Bod J, Della-Giustina D, Tsyrulnik A. Emergency Medicine Residents Consistently Rate Themselves Higher than Attending Assessments on **ACGME Milestones**. West J Emerg Med. 2015 Nov;16(6):931-5.

INTRODUCTION:
In 2012 the Accreditation Council for Graduate Medical Education (ACGME) introduced the Next Accreditation System (NAS), which implemented milestones to assess the competency of residents and fellows. While attending evaluation and feedback is crucial for resident development, perhaps equally important is a resident's self-assessment. If a resident does not accurately self-assess, clinical and professional progress may be compromised. The objective of our study was to compare emergency medicine (EM) resident milestone evaluation by EM faculty with the same resident's self-assessment.
METHODS:

This is an observational, cross-sectional study that was performed at an academic, four-year EM residency program. Twenty-five randomly chosen residents completed milestone self-assessment using eight ACGME sub-competencies deemed by residency leadership as representative of core EM principles. These residents were also evaluated by 20 faculty members. The milestone levels were evaluated on a nine-point scale. We calculated the average difference between resident self-ratings and faculty ratings, and used sample t-tests to determine statistical significance of the difference in scores.

RESULTS:

Eighteen residents evaluated themselves. Each resident was assessed by an average of 16 attendings (min=10, max=20). Residents gave themselves statistically significant higher milestone ratings than attendings did for each sub-competency examined ($p<0.0001$).

CONCLUSION:

Residents over-estimated their abilities in every sub-competency assessed. This underscores the importance of feedback and assessment transparency. More attention needs to be paid to methods by which residency leadership can make residents' self-perception of their clinical ability more congruent with that of their teachers and evaluators. The major limitation of our study is small sample size of both residents and attendings.

5. Friedman KA, Balwan S, Cacace F, Katona K, Sunday S, Chaudhry S. Impact on house staff evaluation scores when changing from a Dreyfus- to a Milestone-based evaluation model: one internal medicine residency program's findings. Med Educ Online. 2014 Nov 24;19:25185.

PURPOSE:

As graduate medical education (GME) moves into the Next Accreditation System (NAS), programs must take a critical look at their current models of evaluation and assess how well they align with reporting outcomes. Our objective was to assess the impact on house staff evaluation scores when transitioning from a Dreyfus-based model of evaluation to a Milestone-based model of evaluation. Milestones are a key component of the NAS.

METHOD:

We analyzed all end of rotation evaluations of house staff completed by faculty for academic years 2010-2011 (pre-Dreyfus model) and 2011-2012 (post-Milestone model) in one large university-based internal medicine residency training program. Main measures included change in PGY-level average score; slope, range, and separation of average scores across all six Accreditation Council for Graduate Medical Education (ACGME) competencies.

RESULTS:

Transitioning from a Dreyfus-based model to a Milestone-based model resulted in a larger separation in the scores between our three post-graduate year classes, a steeper progression of scores in the PGY-1 class, a wider use of the 5-point scale on our global end of rotation evaluation form, and a downward shift in the PGY-1 scores and an upward shift in the PGY-3 scores.

CONCLUSIONS:

For faculty trained in both models of assessment, the Milestone-based model had greater discriminatory ability as evidenced by the larger separation in the scores for all the classes, in particular the PGY-1 class.

6. Hauff SR, Hopson LR, Losman E, Perry MA, Lypson ML, Fischer J, Santen SA. Programmatic assessment of level 1 **milestones** in incoming interns. Acad Emerg Med. 2014 Jun;21(6):694-8.

OBJECTIVES:

With the Accreditation Council for Graduate Medical Education (ACGME) Next Accreditation System, emergency medicine (EM) residency programs will be required to report residents' progress through the EM milestones. The milestones include five progressively advancing skill levels, with Level 1 defining the skill set of a medical school graduate and Level 5, that of an attending physician. The ACGME stresses that multiple forms of assessment should be used to ensure capture of the multifaceted competencies. The objective of this study was to determine the feasibility and results of programmatic assessment of Level 1 milestones using multisource assessments for incoming EM interns in July.

METHODS:

The study population was interns starting in 2012 and 2013. Interns' Level 1 milestone assessment was done with four distinct methods: 1) the postgraduate orientation assessment (POA) by the Graduate Medical Education Office for all incoming interns (this multistation examination covers nine of the EM milestones and includes standardized patient cases, task completion, and computer-based stations); 2) direct observation of patient encounters by core faculty using a milestones-based clinical skills competency checklist; 3) the global monthly assessment at the end of the intern orientation month that was updated to reflect the EM milestones; and 4) faculty assessment during procedural labs. These occurred during the July orientation month that included the POA, clinical shifts, didactic sessions, and procedure labs.

RESULTS:

In the POA, interns were competent in 48% to 93% of the milestones assessed. Overall, competency was 70% to 80%, with low scores noted in aseptic technique (patient care Milestone 13 [PC13]) and written and verbal hand-off (interpersonal communications skills [ICS]2). In overall communication, 70% of interns demonstrated competency. In excess of 80% demonstrated competency in critical values interpretation (PC3), informed consent (PC9), pain assessment (PC11), and geriatric functional assessment (PC3). On direct observation, almost all Level 1 milestones were achieved (93% to 100%); however, only 78% of interns achieved competency in pharmacotherapy (PC5). On global monthly evaluations, all interns met Level 1 milestones.

CONCLUSIONS:

A multisource assessment of EM milestones is feasible and useful to determine Level 1 milestones achievement for incoming interns. A structured assessment program, used in conjunction with more traditional forms of evaluation such as global monthly evaluations and direct observation, is useful for identifying deficits in new trainees and may be able inform the creation of early intervention programs.

7. Warm EJ, Held JD, Hellmann M, Kelleher M, Kinnear B, Lee C, O'Toole JK, Mathis B, Mueller C, Sall D, Tolentino J, Schauer DP. Entrusting Observable Practice Activities and **Milestones** Over the 36 Months of an Internal Medicine **Residency**. Acad Med. 2016 Jun 28. [Epub ahead of print]

PURPOSE:

Competency-based medical education and milestone reporting have led to increased interest in work-based assessments using entrustment over time as an assessment framework. Little is known about data collected from these assessments during residency. This study describes the results of entrustment of discrete work-based skills over 36 months in the University of Cincinnati internal medicine (IM) residency program.

METHOD:

Attending physician and peer/allied health assessors provided entrustment ratings of resident performance on work-based observable practice activities (OPAs) mapped to Accreditation Council for Graduate Medicine Education/American Board of Internal Medicine reporting milestones for IM. These data were translated into milestones data and tracked longitudinally.

The authors analyzed data from this new entrustment system's first 36 months (July 2012-June 2015).

RESULTS:

During the 36-month period, assessors made 364,728 milestone assessments (mapped from OPAs) of 189 residents. Residents received an annualized average of 83 assessment encounters, producing means of 3,987 milestone assessments and 4,325 words of narrative assessment. Mean entrustment ratings (range 1-5) from all assessors for all milestones rose from 2.46 for first-month residents to 3.92 for 36th-month residents ($r = 0.9252$, $P < .001$). Attending physicians' entrustment ratings were lower than peer/allied health assessors' ratings. Medical knowledge and patient care milestones were rated lower than professionalism and interpersonal and communication skills milestones.

CONCLUSIONS:

Entrustment of milestones appears to rise progressively over time, with differences by assessor type, competency, milestone, and resident. Further research is needed to elucidate the validity of these data in promotion, remediation, and reporting decisions.

8. Nabors C, Peterson SJ, Forman L, Stallings GW, Mumtaz A, Sule S, Shah T, Aronow W, Delorenzo L, Chandy D, Lehrman SG, Frishman WH, Holmboe E. Operationalizing the internal medicine milestones-an early status report. J Grad Med Educ. 2013 Mar;5(1):130-7. doi: 10.4300/JGME-D-12-00130.1.

BACKGROUND:

The internal medicine milestones were developed to advance outcomes-based residency training and will play an important role in the next accreditation system.

INNOVATION:

As an element of our program's participation in the internal medicine educational innovations project, we implemented a milestones-based evaluation process in our general medicine and pulmonary-critical care rotations on July 1, 2010.

MEASURES:

Outcomes assessed included survey-rated acceptability to participating faculty, residents, and clinical competency committee members.

RESULTS:

Faculty and residents agreed that the milestones promoted a common understanding of what knowledge, skills, and attitudes should be displayed at particular points in residents' professional development and enhanced evaluators' ability to provide specific performance feedback. Most residents and faculty members agreed that the milestones promoted fairness and uniformity in the evaluation process. Clinical competency committee members agreed the milestones improved the quality of information available for deliberations and resulted in more uniform promotion standards. Faculty rated the use of too many milestones per form/tool at a mean of 7.3 (where 1 was minimally problematic, and 10 was maximally problematic) and the potential for evaluator fatigue (mean, 8.2) as the most significant challenges to the use of milestones. Eight of 12 faculty members would recommend milestones in other programs; 4 were uncertain.

CONCLUSIONS:

Despite logistical challenges, educators and trainees found that milestones promoted a common understanding of what knowledge, skills and attitudes should be displayed at particular stages of training; permitted greater specificity in performance feedback; and enhanced uniformity and fairness in promotion decisions.

Single Institution – Targeted Competency Studies

1. Martin SK, Farnan JM, McConville JF, Arora VM. Piloting a Structured Practice Audit to Assess **ACGME Milestones** in Written Handoff Communication in Internal Medicine. J Grad Med Educ. 2015 Jun;7(2):238-41.

BACKGROUND:
Written communication skills are integral to patient care handoffs. Residency programs require feasible assessment tools that provide timely formative and summative feedback, ideally linked to the Accreditation Council for Graduate Medical Education Milestones.
OBJECTIVE:
We describe the use of 1 such tool-UPDATED-to assess written handoff communication skills in internal medicine interns.
METHODS:
During 2012-2013, the authors piloted a structured practice audit at 1 academic institution to audit written sign-outs completed by 45 interns, using the UPDATED tool, which scores 7 aspects of sign-out communication linked to milestones. Intern sign-outs were audited by trained faculty members throughout the year. Results were incorporated into intern performance reviews and Clinical Competency Committees.
RESULTS:
A total of 136 sign-outs were audited (averaging 3.1 audits per intern). In the first trimester, 14 interns (31%) had satisfactory audit results. Five interns (11%) had critical deficiencies and received immediate feedback, and the remaining 26 (58%) were assigned future audits due to missing audits or unsatisfactory scores. In the second trimester, 21 interns (68%) had satisfactory results, 1 had critical deficiencies, and 9 (29%) required future audits. Nine of the 10 remaining interns in the final trimester had satisfactory audits. Faculty time was estimated at 10 to 15 minutes per sign-out audited.
CONCLUSIONS:
The UPDATED audit is a milestone-based tool that can be used to assess written sign-out communication skills in internal medicine residency programs. Future work is planned to adapt the tool for use by senior supervisory residents to appraise sign-outs in real time.

2. Skinner B, Morgan H, Kobernik E, Kamdar N, Curran D, Marzano D, Hammoud M. The Decision to Incision Curriculum: Teaching Preoperative Skills and Achieving Level 1 **Milestones**. J Surg Educ. 2016 Jul-Aug;73(4):735-40

OBJECTIVE:
To evaluate the effectiveness of a preoperative skills curriculum, and to assess and document competence in associated Obstetrics and Gynecology Level 1 Milestones.
DESIGN:
The Decision to Incision curriculum was developed by a team of medical educators with the goal of teaching and evaluating 5 skills pertinent to Milestone 1: Preoperative consent, patient positioning, Foley catheter placement, surgical scrub, and preoperative time-out. Competence, overall skill performance, and knowledge were assessed by evaluator rating using checklists before and after the educational intervention. Differences between preintervention and postintervention skills performance and competence were assessed using Wilcoxon rank test and Fisher exact test, respectively.
SETTING:
Clinical Simulation Center at an academic medical center.
PARTICIPANTS:

Overall, 29 fourth year medical students matriculating into Obstetrics and Gynecology residencies.
RESULTS:
The proportion of participants meeting Milestone competence significantly increased in all 5 skills, with competence achieved in 95.6% (95% CI: 92.1-99.0) of posttest skills assessments. Median overall performance also significantly improved for all 5 skills, with 83.6% (95% CI: 77.3-89.9) earning scores of 4 out of 5 or greater on the posttest. For knowledge testing, the proportion of correct responses significantly increased for both topics evaluated, from 45.2% to 99.7% ($p < 0.0001$) for positioning and from 32.8% to 83.1% ($p < 0.0001$) for time-out.
CONCLUSIONS:
The decision to incision curriculum significantly improved preoperative skills, including skills that may be required on day 1 of residency. This curriculum also facilitated achievement and documentation of competence in multiple Milestones.

3. Safir IJ, Shrewsberry AB, Issa IM, Ogan K, Ritenour CW, Sullivan J, Issa MM. Impact of remote monitoring and supervision on resident training using new **ACGME** milestone criteria. Can J Urol. 2015 Oct;22(5):7959-64.

INTRODUCTION:
The study objective was to determine the impact of remote monitoring and supervision (RMS) in integrated endourology suites (IES) on residents achieving endoscopic training milestones.
MATERIALS AND METHODS:
Twenty-one urology residents evaluated RMS in IES using a 25-question survey. IES provided audio-visual communication for faculty to supervise residents remotely. Questionnaire used a linear visual scale of 1-10 to assess acceptability (8 questions), impact on training (10 questions), supervision level (1 question), and pre- and post-training milestone self-assessments (6 questions). Improvements in Patient Care Milestone #7 (upper/lower tract endoscopic procedures) and Patient Care Milestone #9 (office-based procedures) were analyzed.
RESULTS:
Twenty-one urology residents (out of potential 23) evaluated RMS in IES using a 25-question survey (91.3% response rate). Overall RMS acceptability and satisfaction was high (mean score = 9.1/10) with a majority (95.2%) feeling comfortable being alone with the patient. Residents reported positively on the following parameters: autonomy without compromising safety (8.7), supervision level (8.6), achieving independence (8.4), education quality (8.3), learning rate (8.1), clinical decision-making (8.0), and reducing case numbers to achieve proficiency (7.6). Residents perceived no issues with under- or over-supervision, and a majority (76.2%) expressed that RMS should be standard of training in residency programs. Residents reported mean level increases of 2.5 and 2.8 (out of 5) in Patient Care Milestones for endoscopic procedures and office-based procedures, respectively ($p < 0.0001$).
CONCLUSIONS:
RMS in integrated endourology suites may enhance resident education and endoscopic training. The study demonstrated an increase in competency levels reported by residents trained using RMS.

4. Sanford B, Whitehouse S, Kokas M.Do **ACGME** Physician-Patient Communication **Milestones** Align With HCAHPS Patient Satisfaction Measures for Doctor Communication? Am J Med Qual. 2016 May 27. pii: 1062860616652938. [Epub ahead of print]

-no Abstract available.