Title
Essential operating principles of cellular morphogenesis

Author
Engelberg, Jesse

Publication Date
2011

Peer reviewed|Thesis/dissertation

Essential Operating Principles of Cellular Morphogenesis

by

Jesse Aaron Engelberg

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

[redacted]

For my family

You are the best

Acknowledgements

Thank you. Thank you to my family and to my friends. I am lucky to have you as a part of my life, and indebted to you for your support. Thank you Julia for inspiring me to finish this part of my life and move into the next part with you. Thank you Virginia for keeping me sane during this process. Thank you Mom, Dad, Kyra, Ari, Kellie, Mikaela, Saghi, Naz, and Demetrius for being a wonderful family and always making me feel welcome and cared for. Thank you roommates, neighbors, teammates, and all the wonderful friends who supported me and made this worth it.

Thank you to my colleagues in the Hunt Lab and the Mostov Lab. Thank you Keith, for providing me with a forum to learn more about cell biology and to explore the new scientific frontiers I was struggling with. Thank you Rana for your collaboration on the MDCK project and being willing to listen and explain how and why things worked. Thank you Donna, Keith, Zena, and Tony for reading this dissertation and providing feedback and advice during the process.

Thank you, Tony, for guiding me and advising me along the way. You helped me find direction when I wasn't sure where to go, helped me plan, write, and edit. I am proud of what we've accomplished together, and glad to have been your student.

Contributions of others to the presented work

The content of this manuscript is primarily derived from two manuscripts. The first, "Essential operating principles of tumor spheroid growth" was published in *BMC Systems Biology* in December 2008. The second, "MDCK cystogenesis driven by cell stabilization within computational analogues" is in press at *PLoS Computational Biology*. The introduction and summary contain some text from both papers, while Chapter 2 is directly adapted from the *BMC Systems Biology* paper and Chapter 3 from the *PLoS Computational Biology* paper. The in vitro and in silico data described in Chapter 4 are original, but the structure parallels that of Chapter 3.

I wrote the text of both papers, with additional writing, editing, and supervision by my advisor, C. Anthony Hunt. Glen E.P. Ropella contributed to the writing and editing of the *BMC Systems Biology* paper, while Anirban Datta and Keith Mostov contributed written portions and edited the *PLoS Computational Biology* paper.

Glen E.P. Ropella contributed to the computational research underlying Chapter 2, while Anirban Datta performed the in vitro experiments described in Chapters 3 and 4. The in silico analogue detailed in Chapters 3 and 4 was inspired by an earlier analogue developed by Mark Grant. Appendix 2 contains the text of a paper in press at the 2011 Annual Agent-Directed Simulation Symposium.

Essential operating principles of cellular morphogenesis

by

Jesse Aaron Engelberg

C. Anthony Hunt
Committee Chair

Abstract

The study of epithelial morphogenesis is fundamental to increasing our understanding of organ function and disease. Great progress has been made through study of culture systems such as EMT6/Ro mouse mammary tumor spheroids and Madin-Darby canine kidney (MDCK) cells, but many aspects of even simple morphogenesis remain unclear. For example, are specific cell actions tightly coupled to the characteristics cell environment or are they more often cell state dependent? Our objective was to discover plausible representations of the operating principles realized during characteristic growth of EMT6/Ro mouse mammary tumor spheroids and MDCK cells in culture. To reach that objective we engineered and iteratively falsified agent-oriented analogues of EMT6 spheroid growth (Chapter 2) and MDCK cystogenesis (Chapters 3 and 4). An approach to understanding how individual cell behaviors contribute to system behaviors is to discover a set of principles that enable abstract agents to exhibit closely analogous behaviors using only information available in an agent's immediate environment. EMT6 spheroids and MDCK cysts display consistent and predictable growth characteristics, implying that individual cell behaviors are tightly controlled and regulated. We listed key attributes of EMT6 spheroid growth and MDCK cystogenesis, which became our

behavioral targets.

To understand the in vitro systems we created analogues made up of quasi-autonomous software agents and an abstract environment in which they could operate. The EMT6 analogue was designed so that upon execution it could mimic EMT6 cells forming spheroids in culture. Each agent used an identical set of axiomatic operating principles. In sequence, we used the list of targeted attributes to falsify and revise these axioms, until the analogue exhibited behaviors and attributes that were within prespecified ranges of those targeted, thereby achieving a level of validation. For the analogue of MDCK cystogenesis we tested our hypotheses through in vitro experimentation and quantitative validation. We observed novel growth patterns, including a cell behavior shift that began around day five of growth.

We posit that the validated analogues' operating principles are reasonable representations of those utilized by EMT6/Ro and MDCK cells during morphogenesis. Simulations provide an observable theory for cystogenesis based on hypothesized, cell-level operating principles.

Table of Contents

List of Tables

List of Figures

1. Introduction

Extensive study of epithelial cells grown in culture has provided useful insight into important aspects of cell growth and tissue organization. The behaviors of epithelial cells in culture fall reliably within narrow ranges, as if cell behavior and thus the underlying mechanisms are tightly choreographed. Those actions can be thought of as being constrained and guided by a set of genetically specified biological operating principles. Can we discover and attribute a small, robust set of operating principles that combine to create the system level phenomena that characterize epithelial cell growth in vitro? How can we represent and challenge those operating principles? What organization of the subcellular molecular biology enables the operating principles to emerge, and be sustained at the cellular level? Before addressing the last question, we need answers to the first two, which has been the objective of the projects described within this document. We created two distinct analogues of epithelial cell behavior and organization. The first, specified in Chapter 2, was an analogue of in vitro multicellular tumor spheroid growth. The second, specified in Chapters 3 and 4, was an analogue of MDCK cell cytogenesis within embedded Matrigel or collagen culture.

An approach to understanding how individual cell behaviors can contribute to a diverse set of system level attributes is to discover a set of simple yet sufficient principles that enable abstract, cell mimetic agents, using only locally available information, to exhibit behaviors closely analogous to cells in culture. For this context, we define a biological operating principle to be an abstract, inferential representation of an action within a reliably behaved cell system. To discover these principles, we created quasi-autonomous computer analogues comprised of individual cell mimetic agents (CELLS)

that adhered to a common, small set of axiomatic operating principles. We use axiom as commonly defined [1] and to emphasize that the analogue, unlike epithelial cells in culture, is a formal mathematical system and its execution is a form of deduction from the axioms within the analogue. Hereafter, we use AXIOM to emphasize that we refer only to the computational analogues. An AXIOM specified a behavior that depended on the local environment perceived by the CELL, given its internal state. Individual AXIOMS were implementations of in silico, axiomatic operating principles. Each axiomatic operating principle was derived from a postulated in vitro counterpart as described in the Methods sections. The combined actions of an expanding population of CELLS, each adhering to the same set of operating principles, were sufficient to produce unique systemic behaviors. The systems underwent several rounds of iterative refinement and parameter tuning. When measured, the resulting behaviors provided a set of systemic attributes that matched observed in vitro attributes closely for both systems. Once that was achieved, we could postulate that the axiomatic operating principles may have in vitro counterparts, as illustrated in Figure 1.1. To date, such principles have been arrived at piecemeal by induction following experimentation. Experimental cell biology has been successful at discovering isolated cell level operating principles, but progress has been slow in providing a unified understanding of autonomous cellular behavior. We anticipated that iterative analogue refinement would lead to improved insight into cell level operating principles and plausible mappings to their biological counterparts.

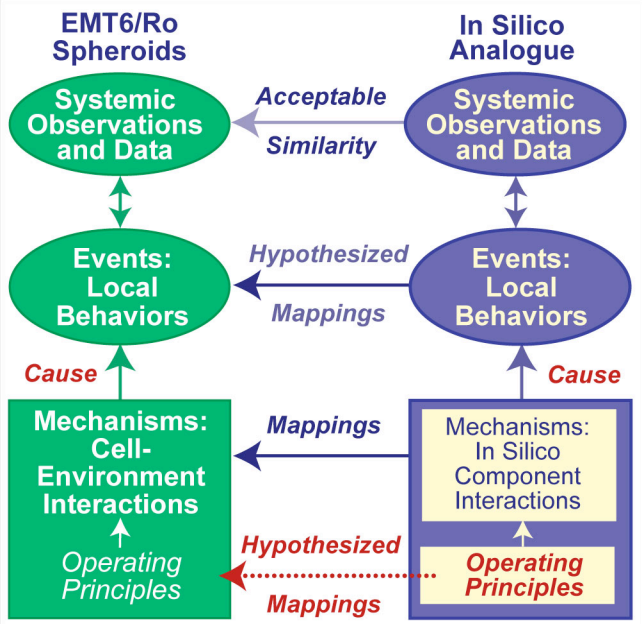


Figure 1.1. Relationships between simulated multicellular tumor spheroids (SMS) and EMT6 spheroids.
An SMS is comprised of quasi-autonomous cell components interacting with adjacent cells and factors in their environment by adhering to a set of axiomatic operating principles. A clear mapping exists between SMS components and EMT6 counterparts. Following execution, the interacting components cause local and systemic behaviors. Measures of cell and system behaviors provide a set of attributes—the SMS phenotype. Validation was achieved when SMS attributes were measurably similar to a targeted set of EMT6 attributes. When that was accomplished, we could hypothesize that a semiquantitative mapping exists between in silico and in vitro events. We could also hypothesize that the set of axiomatic operating principles has a biological counterpart.

A cohesive set of operating principles (as distinct from isolated principles) can provide a framework into which more detailed, subcellular and molecular level information can be connected directly to system level phenotype. The plan was to work backward from a targeted set of in vitro observations of cell and system level phenomena to a plausible set of analogue AXIOMS, which would be necessary and sufficient to generate in silico counterparts of the targeted phenomena. With that vision, these projects have been motivated by three expectations: 1) Understanding hypothesized mechanisms in vitro would be facilitated by successfully building and studying analogous mechanisms in silico. 2) Achieving and refining validated analogues would offer a scientific, experimental approach to discovering and studying cohesive sets of operating principles.

3) Knowledge of axiomatic operating principles would facilitate exploration of their biological counterparts. This dissertation reports on the design and implementation of two analogues, and the results of their execution.

1.1. Previous modeling efforts

1.1.1. Tumor spheroid modeling

Efforts to model tumor spheroid growth characteristics have been extensive, informative, and successful (see [2], two recent reviews [3, 4], and references therein). However, no one has proposed a cohesive set of cell level operating principles. Only recently, due to advances in computer processing power and advance software modeling techniques, has it become feasible to design and instantiate quasi-autonomous, cell mimetic analogues, [5-7] capable of exhibiting a rich phenotype of their own. The focus of most modeling and simulation efforts has not been in that direction. It has been primarily to provide precise, mostly mathematical descriptions of growth dynamics in terms of measured biochemical and physical factors combined with detailed descriptions of essential cell processes [8-12]. Casciari et al. expanded on previous work to produce a differential equation-based model that addressed the effect of nutrient and ionic diffusion on tumor spheroid growth. The model specifically considered the gradients in the concentration of oxygen, glucose, lactation, carbon dioxide and hydrogen ions. It was hypothesized that the gradients in oxygen, glucose, and pH contribute to the heterogeneity of EMT6 growth. The model successfully predicted EMT6 growth rates at under certain conditions, but could not predict the onset of necrosis or the stabilization is EMT6 size.

Detailed modeling of the metabolic processes underlying tumor spheroid growth

has also been attempted. Chignola et al. [10] developed a computational model to study tumor spheroid growth that involved a number of Michaelis-Menten-like equations describing glucose metabolism. The model includes parameters that describe the activity of glucokinase, glucose transporters, and glucose phosphorylation, among others. In addition they modeled the cell cycle of tumor cells in such a way that it depends on cell metabolism. Various energy checkpoints must be met in order for cells to proliferate, slowing growth within the inner layers of the spheroid. In order to validate the model simulations were run with individual simulated cells. Researchers were able to mimic aspects of in vitro metabolism including changes in the rate of oxygen consumption when environmental glucose levels were changed and the relationship between the cell cycle and levels of ATP.

The resulting models have been successful in explaining the time course and limits of spheroid growth in terms of nutrient depletion [8], increased acidity near the spheroid's center [9], and the dynamics of tumor spheroid metabolism [10]. Jiang et al. combined these features into a comprehensive model that separately considered each cell and spanned three levels of mechanistic resolution [2]. Other modeling efforts such as [11, 12] have used hybrid mathematical and individual based approaches that have shown initially promising qualitative results.

1.1.2. Modeling epithelial cell cystogenesis

There are a number of existing analogues of epithelial cystogenesis, but only two specifically model MDCK cell cytogenesis. Single cell-based models of epithelial morphogenesis are reviewed in [13]. More recently, Rejniak developed a model of epithelial cell cystogenesis [14] that mimicked the growth of MCF-10A cells in culture.

It could be argued that [14] could be adapted to mimic aspects of MDCK cell behavior, but to do so would require carefully revalidating the model against phenotypic data from MDCK epithelial cystogenesis. The underlying assumptions that led to the creation of the model would have to be rechecked, and any differences in MCF-10A and MDCK cell growth would have to be addressed. In addition, a new set of targeted attributes and measures of similarity would have to be assembled and the model revalidated against these different lists. If data used to create the original model were not available for MDCK cell cystogenesis that data would have to be produced through in vitro experimentation.

The existing models of MDCK cystogenesis include the work by Grant et al. [5] and the updated version [15]. Both of these works used a similar approach as the current work, but there were a number of significant differences in the implementation. Briefly, Grant et al. constructed their model using Mason (http://cs.gmu.edu/~eclab/projects/mason), an agent-based modeling software technology on a hexagonal grid with polarized and unpolarized cells, luminal space, and extra cellular matrix. Each cell, luminal space, or matrix took up exactly one space within the grid. The model used axioms to dictate cell behavior, but these axioms were entirely controlled by the neighborhood in which the cell resided. The list of targeted attributes was primarily qualitative, based on mimicking MDCK cell behavior in a number of different culture environments, including embedded culture, suspension culture, overlay, and sandwhich culture. The analogue was the first to successfully mimic MDCK cell behavior in these different culture conditions. In a follow-on work, Kim et al. [15] modified the model so that each axiom could be applied probabilistically and then tested

the results of relaxing certain axioms to measure in silico cell behavior when axioms were not followed stringently. Significantly, they discovered that certain behaviors, such as the orientation of cell division, could be relaxed without affecting cyst growth in a qualitative manner, but others, like simulated anoikis, would dramatically alter cyst growth when not followed. Although successful in many ways, these analogues had a number of drawbacks. One example is that they could not mimic the commonly observed multiple lumen phenotype. In addition the models were not validated against quantitative data. The models would saturate in cyst size and area relatively quickly and not continue to proliferate. Also, because each cell only occupied a single location on the grid, cell shape and size played no part in the model and cell division was not strictly biological.

1.2. Executable cell biology

The analogues described herein are examples of a class of simulation models referred to as executable biology [16, 17]. Our simulations share some similarities with [18] and are closely related to [19], though the systems under study and the simulation frameworks are distinct. Executable biological analogues are poorly suited for precise prediction, but are ideally suited for testing hypotheses about mechanisms. The basic method requires building mechanisms at the functional unit level closest to the targeted phenomena. Here, that unit is the cell. Our analogues are comprised of quasi-autonomous agents. Each maps to an EMT6 or MDCK cell. The initial limit for SMS resolution was the CELL, while the MDCK analogue used a more detailed level of resolution that allowed for shape change and more realistic CELL division. If achieving

the list of targeted attributes required doing so, the resolution of the SMS could be increased.

1.3. Iterative Refinement Process

Once a set of targeted attributes had been specified, parameter tuning and analogue validation became closely linked. When seeking fundamental necessary and sufficient in silico mechanisms, we incremented the complexity upwards. We started with the simplest possible system and used an iterative falsification process, beginning with the first of the targeted attributes listed in Table 1.1 (for the SMS) and Table 2.2 (for the ISMA). That iterative refinement method, of which parsimony is a factor, has been used successfully in addressing other simulation goals [20]. For the SMS, while exploring early AXIOM specifications and the in silico conditions needed to achieve the first attribute, we mostly ignored our larger knowledge of EMT6 spheroid biology. At that stage, the analogue had one and only one goal: achieve the targeted attribute. Once that was achieved, that early SMS was valid *for that one targeted attribute.* In this context, an analogue was considered valid if it exhibited attributes that matched the targeted set according to some prespecified similarity measure. We executed that same protocol for each of the other attributes in Table 1.1. Following each expansion of the attribute list, we reconsidered all AXIOMS, revising and merging parsimoniously where needed. We initially coarse-tuned parameter values and subsequently fine-tuned them. Our explicit process of iterative falsification contrasts to most prior work in executable biology, including [18, 19], which describe the completed models and predictions, but do not list the attributes targeted for reproduction or the order in which they were achieved. We believe added transparency will allow others to build on the work described here.

Table 1.1. Targeted attributes.

Attribute	Description
1	Cells consume resources, change state, proliferate, lose adhesion, die, shed, and move.
2	Cells proliferate throughout the duration of growth of the EMT6 spheroid.
3	Cells behave autonomously and locally.
4	The EMT6 spheroid develops an inner necrotic core, a middle quiescent layer, and an outer proliferating layer.
5	The EMT6 spheroid initially grows exponentially, then linearly, and then stabilizes.
6	The EMT6 spheroid has different growth characteristics at different levels of nutrient.
7	Necrosis onset occurs when the EMT6 spheroid has an area of roughly 0.2 mm^2 at high nutrient and 0.02 mm^2 at low nutrient.
8	The viable rim has a width of roughly 240 μm at high nutrient and 60 μm at low nutrient.
9	The measured initial doubling times are roughly 22 hours at high nutrient and 26 hours at low nutrient.
10	The mean error percentage between EMT6 spheroid and SMS growth is within 15% at high and low nutrient levels.

Attributes of in vitro growth targeted during development and refinement of the SMS.

1.4. Summary

The work described within this dissertation documents an evolving process. The model of tumor spheroid growth described a complex process in a straightforward manner, successfully increasing our knowledge of the underlying biological system. As new knowledge was gained, it became clear that continued development of the model would require increased in vitro data, not available at the time of execution. The existing collaboration established during the preparation of [5] (which modeled MDCK cystogenesis) provided a good starting point for the development and iterative refinement of a new analogue, however. Expanding the targeted attribute list for MDCK cystogenesis to encompass cell size and shape change quickly falsified the earlier analogue, requiring the development of the new analogue described in Chapter 3. Through a cooperative process we executed in silico and in vitro experiments that

allowed the ISMA to closely mimic data from in vitro experiments. Further knowledge was gained when the ISMA was expanded to encompass results from MDCK cystogenesis within collagen, as discussed in Chapter 4. The steady improvement of the underlying method of refining in silico analogues allowed us to gain an increased understanding of analogue development and of the events that occur during epithelial cell morphogenesis.

2. Essential operating principles of tumor spheroid growth

2.1. Background

The objective of this project has been different: we aimed to discover a somewhat minimalist set of essential axiomatic operating principles that would enable the system level growth characteristics produced by CELLS to match a targeted set of tumor spheroid growth attributes, based on measures of similarity. Given that constraint, we identified nine axiomatic operating principles. To narrow the scope and to achieve one of the key targeted attributes, we insisted that CELLS only acquire and use information available locally. We designed the system so that systemic properties were a consequence of only local CELL interactions. We made it easy to revise CELL AXIOMS so that we could sequentially expand the set of targeted attributes achieved. The targeted attributes achieved (Table 1.1) include those that are most characteristic of in vitro tumor spheroid growth: development of a stable size; a three-layered structure that maps to outer, proliferating cells and a core of inner, necrotic cells, with quiescent cells in between; the ability of cells to shed; and realistic growth curves under two different growth conditions. To achieve the targeted attributes, it was not necessary to simulate the release of growth inhibitory substances. Our objective is to understand the behavior of early-stage avascular cancers, and as a result we chose to model the EMT6 cell line. Though it is tumor-derived it has proven stable and exhibits reliable behaviors. For the purposes of this research, we can treat EMT6 cells as being in a healthy, not diseased state.

2.1.1. In vitro system: historical context

Because cancer is such a complex and heterogeneous disease, researchers develop

and study model systems. One is the in vitro EMT6/Ro multicellular tumor spheroid system. Freyer and Sutherland used the system to study avascular cancers in the 1970s and 1980s [21-26]. Their initial goal was to create a system that would allow many EMT6 spheroids to be grown in the same flask under identical, controlled conditions. Study of that model was expected to improve our understanding of how early stage cancer forms and improve our ability to treat it when the cells have not reached total genomic instability and still have much in common with normal cells [27].

In order to obtain adequate numbers of spheroids for measuring growth curves, experiments employed spinner flasks containing hundreds of spheroids. The cultures were initiated in monolayer and then grown in dishes until small spheroids were present (95-100 hours). These spheroids (usually 400-600 cells) were sorted and transferred to flasks, which contained a solution of glucose-free Eagle's Basal Medium, Fetal Bovine Serum, and varied concentrations of glucose. Oxygen was bubbled through the flask, and glucose was replenished roughly every 10-14 hours [21]. Most early experiments were designed to characterize the system and its behavior. Eventually, however, many researchers shifted to using the system as a tool rather than studying the system itself. The seminal studies describing the behavior and characteristics of EMT6 spheroids were primarily completed by 1992. Important work continued nevertheless. There was an effort [28] to identify a potential necrotic inhibitor, and confocal microscopy was used to assess growth fractions [29].

2.2. Methods

2.2.1. Targeted attributes describe in vitro EMT6 spheroid growth

Although there is variation in EMT6 spheroids' growth, it reliably follows the

same well-defined pattern [23]. EMT6 spheroids initially grow exponentially without constraint from nutrient or other cells. This gives way to linear growth as cells become quiescent due to nutrient depletion within the spheroids. Eventually the spheroids begin stabilizing, both in volume and cell number, though cells continue to reproduce on the outer edge. They develop a concentric layered structure: an outer layer of actively proliferating cells, a middle layer of quiescent cells, and a core of necrotic cells and cellular debris. The material released by dying cells is thought to inhibit cellular proliferation [30], but it is not clear if this material actually affects cells in the rim and thus spheroid growth rates. During growth, an EMT6 spheroid maintains a generally spherical shape, but neither shape nor relative EMT6 spheroid stability have been quantified, which makes shape validation difficult. Additionally, though the width of the viable rim has been measured for different cell types, only one group has attempted to quantify the ratio of proliferating and quiescent cells [29]. Wartenburg and Acker performed these measurements on human glioma spheroids. They do form spheroids with concentric layers, but have quantitative growth characteristics (such as initial doubling time and saturation size) that are different from those of EMT6 spheroids. We elected to falsify and validate our experimental results using data from Freyer and Sutherland [21], because they performed the most thorough and complete analysis of EMT6 spheroid growth, including measuring growth curves, the width of the viable rim, and the size at the approximate time of necrosis onset.

2.2.2. Analogue construction within an agent-based paradigm

SMS construction used agent-based methods available in the simulation toolkit MASON [31]. This framework was used for data generation, scheduling, and

visualization. To distinguish clearly in silico components and processes from corresponding components and processes within EMT6, we use SMALL CAPS when referring to the former. Variable names are in *italics*, and each is defined in the order it is introduced. CELLS interact with each other and their environment during each simulation cycle within a two-dimensional, hexagonal grid. Earlier versions used a square grid, but in addition to requiring a higher order implementation of discrete diffusion, it also generated artifacts that are not present with the hexagonal grid. We tested different orders of discrete diffusion to verify that the diffusion algorithm did not cause artifacts. CELL actions are mandated by AXIOMS [5, 7]: when a specified condition is met, a specified action occurs. Together, these AXIOMS are a CELL's operating principles. A goal has been to improve the variety of SMS attributes that are similar to corresponding EMT6 spheroid attributes: the expectation being that with increasing phenotypic similarity, the higher the likelihood SMS AXIOMS will map to corresponding EMT6 operating principles (Figure 1.1). As the list of targeted attributes expands, analogue resolution can be adjusted as needed.

A second grid, adjacent to the CELL'S grid contains a diffusible substance called NUTRIENT. NUTRIENT adjacent to each CELL is detected by and available to that CELL for consumption. During each simulation cycle, each CELL uses AXIOMS to select actions based on how its local environment has changed since the last simulation cycle. Examples of actions include move, change state, create a new CELL, DIE, and shed. AXIOMS are implemented by algorithms that utilize the parameter values listed in Table 2.1. AXIOMS are listed in Table 2.2. Where appropriate, parameter values intentionally mirror values measured in vitro. The remaining parameters were tuned using an iterative

process: change parameter value, execute, evaluate relative to referent observables, cogitate, change again, etc.

Table 2.1. Parameter names, values, units and sources.

Parameter	In silico value	In vitro value	Source
Proliferating NUTRIENT critical level (*proNut*)	3.0×10^{-3}	3.0×10^{-19} mol / μm^3	Tuned parameter
Quiescent NUTRIENT critical level (*quiNut*)	8.0×10^{-4}	8.0×10^{-20} mol / μm^3	Tuned parameter
Proliferating CELL'S NUTRIENT uptake (*proConsumeRate*)	5.0×10^{-4} / SEC	5.0×10^{-17} mol / (cell s)	{{;50 Casciari,J.J. 1992 }}]
Quiescent CELL'S NUTRIENT uptake (*quiConsumeRate*)	1.0×10^{-4} / SEC	1.0×10^{-17} mol / (cell s)	Tuned parameter
Delay before dead CELL is removed (*removeDelay*)	3.6×10^{4} SEC	1.8×10^{4} s	{{;74 Harris,L.K. 2006 }}
Movement bias (*moveEmptyBias*)	1.0	—	Tuned parameter
Delay between CELL creation events (*prolifDelay*)	800 SEC	800 s	Tuned parameter
Proliferation bias (*proBias*)	2.25	—	Tuned parameter
NUTRIENT diffusivity (*diffusionRate*)	0.28*	105 μm^2 / s	{{;50 Casciari,J.J. 1992 }}
Initial NUTRIENT concentration (*initialVal*)	0.165 or 0.008	16.5 mM or 0.8 mM	[26]
Time step	1.0 SEC	1.0 s	Calculated
CELL diameter	1 grid space	10 µm	[25]
Average cell cycle**	~4.24×10^{4} SEC	~4.24×10^{4} s	Calculated
Average time of removal after cell death**	1.8×10^{4} SEC	1.8×10^{4} s	Calculated

Sourced parameters were obtained from literature. Tuned parameters were adjusted to allow the analogue to mimic in vitro data. Calculated parameters depended on sourced or tuned parameters. All parameters were fixed, except *initialVal*, which was set to high (equivalent to 0.28 mM oxygen and 16.5 mM glucose) or low (equivalent to 0.07 mM oxygen and 0.8 mM glucose) in different simulations. *: The value of *diffusionRate* is related to the simulation cycle and CELL diameter, as described within the text. **: The mean value observed during simulation.

Table 2.2. SMS axioms.

AXIOM	Environment	Action	Parameters used	In vitro source
1	NUTRIENT > *proNut*	Switch to PROLIFERATING state	*proNut*	Cell quiescence is regulated by the glucose and oxygen supply [32].
2	NUTRIENT < *quiNut*	Switch to NECROTIC state	*quiNut*	Cell death is regulated by the glucose and oxygen supply [32].
3	*quiNut* < NUTRIENT < *proNut*	Switch to QUIESCENT state	*proNut, quiNut*	Cell quiescence is regulated by the glucose and oxygen supply [32].
4	State = PROLIFERATING or QUIESCENT	Consume NUTRIENT equal to *proConsumeRate* or *quiConsumeRate*	*proConsumeRate, quiConsumeRate*	Cells consume oxygen and glucose at varied levels {{;50 Casciari,J.J. 1992 }}.
5	State = NECROTIC; *removeCounter* < 0	Remove CELL	*removeDelay*	Necrotic cells eventually break up and are consumed [33].
6	Inside CELL adjacent to empty space	Move into empty space		Cells move and mix with other cells within the spheroid [34].
7	Outside CELL adjacent to empty space	Move into empty space with prob. p_m	*moveEmptyBias*	Cells move and mix with other cells within the spheroid [34].
8	Outside CELL with 0 neighbors	Randomly move in space		Cells can be shed from the exterior of the spheroid [35].
9	State = PROLIFERATING; *prolifCounter* < 0; CELL has empty neighbors	Create new CELL with prob. p_b	*prolifDelay, proBias*	Cells create new cells within the SMS, causing it to increase in size [32].

In each time step a CELL will execute one or more of these AXIOMS based on the environment the CELL is exposed to and its internal variables. Random numbers for probability functions are generated from a uniform distribution in [0,1), except as noted in the text.

2.2.3. Tuning parameter values improves the analogue's ability to survive falsification

To achieve the specified targeted attributes, we used the IR Protocol, beginning with the first of the targeted attributes listed in Table 1.1. As shown in Table 2.2, during the process of achieving Attributes 1 and 2 individual AXIOMS were qualitatively validated against their in vitro counterparts. For instance, we verified that individual CELLS did not create new CELLS more frequently than is observed in vitro. Once the SMS was validated for that attribute, we added a new attribute, such as no. 2 in Table 1.1, to the targeted list. Doing so often (but not always) immediately falsified that SMS, which was the case with the addition of attribute no. 2.

To revise the construct and form a new, more valid SMS with the expanded set of targeted attributes, we found it essential to introduce a volume loss mechanism and a

mechanism to stabilize SMS growth and shape: we added CELL shedding and STRESS states along with AXIOMS to manage those new states. The new AXIOMS necessitated adding new parameters: *moveEmptyBias and proBias*. Following a period of iterative refinement, these additional mechanisms enabled the SMS to survive our attempts to falsify it with the expanded attribute list. We continued that process for all attributes listed in Table 1.1, until the SMS was able successfully mimic growth curves from EMT6 growth. That same iterative refinement method can be used to further improve SMS behaviors, and—presumably—bring SMS principles of operation into closer alignment with those of EMT6 cells.

Some in silico parameters, such as the *diameter* of a CELL, mapped directly to measured observations of in vitro EMT6 quantities. These are noted by their source in literature within Table 2.1. One exception was the mean value of the in silico interval between when a CELL entered the NECROTIC state and when it was removed (creating an empty space). A value of five HOURS (18,000 SECONDS) was used. Doing so required three assumptions. The first was that the experimental setup used by Harris et al. to obtain these measurements did not contribute excessively to the measured apoptosis duration [33]. The second was acceptance of the authors' assumptions about apoptosis: apoptosis begins when apoptotic morphology was observed and ended when the cell began to fragment. The final and most significant was that we could map these values to EMT6 cells undergoing necrotic cell death induced by nutrient depletion.

The actual number of simulation cycles that elapsed from when a CELL became NECROTIC and when it disappeared depended on the value of *removeCounter*, a pseudo-random number (PRN) drawn from a uniform distribution over the interval [0–

removeDelay). *RemoveCounter* was decremented each cycle that a CELL was in the NECROTIC state, resulting in its removal when the value reached 0. Setting *removeDelay* to 36,000 SECONDS (SEC), resulted in a mean *removeCounter* = 18,000 SEC, which mapped directly to the reported mean duration of apoptosis [33].

In order to achieve the targeted attributes, it was sometimes necessary to select parameters that mapped to values that were toward the extreme end of an observed, referent range. For instance, in order to avoid excessive NUTRIENT consumption resulting in premature appearance of NECROSIS, CELLS consumed NUTRIENT at a rate of 5.0×10^{-17} MOL/CELL/SEC. Observed glucose consumption rates were between 5.5×10^{-17} and 36.0×10^{-17} mol/cell/s [36].

Once a subset of parameter values had been set to map to in vitro counterparts, the remaining parameter values were tuned empirically so that the similarity between SMS and in vitro attributes achieved a specified measure of similarity. Previous agent-based simulation projects demonstrated that the empirical tuning approach is an effective strategy for locating biologically relevant regions of an analogue's parameter space [5, 7, 33]. Initially, parameter values were varied extensively to discover ranges for which qualitative SMS behavior could be mapped to a corresponding biologically plausible behavior. For instance, if *prolifNut* (the value that must be exceeded for a CELL to remain in the PROLIFERATING state) was higher than *initialVal*, proliferation did not occur. Similarly, we found that *moveEmptyBias* had to be higher than 0.5 to prevent fissure formation and eventual SMS destabilization. Following empirical tuning to the in vitro doubling times of 18 to 24 hours [21], we selected a value of 2.25 for *proBias* and 800 SEC for *prolifDelay*. Each CELL had its own individual *prolifCounter* that specified the

number of SEC that must pass before it attempted to create a new CELL. We calculated *prolifCounter* using the method in Walker et al. [37]: *prolifCounter* = *prolifDelay*/2 + R_G, where R_G was a pseudo-random number drawn from a Gaussian distribution having mean = *prolifDelay*/2 and standard deviation = *prolifDelay*/10 [37]. Consequently, the average *prolifCounter* value was roughly equal to *prolifDelay*. Once parameter ranges were identified that achieved the targeted measure of similarity, each parameter was adjusted in sequence over a narrow range, and the consequences for SMS properties were recorded. Values that brought simulated behaviors closer to targeted values were retained. The parameter values obtained following that protocol are identified in Table 2.1. Note that with the possible exception of the critical NUTRIENT levels, none of these tuned parameter values map directly to measurable in vitro counterparts, and it would be problematic to obtain such values through experimentation.

2.2.4. Measuring in silico and in vitro values

In vitro doubling time is the time required for an average EMT6 cluster to grow from 600 to 1,200 cells [21]. These numbers corresponded to an SMS expanding from 8.6×10^{-3} mm^2 to 1.35×10^{-2} mm^2, which we used to determine SMS doubling time. Both high and low NUTRIENT VIABLE rim values were calculated by averaging the VIABLE rim width at NECROSIS onset and at the end of the simulation. Individual values of VIABLE rim width were found by counting the number of CELLS between the SMS center and the edge in three directions: from right of center, above the center, and diagonally left of the center. These three values were averaged to obtain the final width. Because the initial occurrence of NECROSIS was not necessarily stable (during early growth NECROSIS could appear and later vanish), we estimated the time of NECROSIS onset by moving backward;

we specified it to be the latest time at which no NECROTIC CELLS were present. The method is similar to the one used by wet-lab researchers to estimate necrosis onset [21]. As done with EMT6 spheroids, the maximum SMS size reached was estimated by fitting the Gompertz equation to growth data and taking the maximum size predicted by the equation.

Growth rates in vitro were based on mean measures of spheroid diameter [21]. We converted those values into cross-sectional area in order to compare them with our in silico results. To establish a measure of SMS cross-sectional area we followed the method used in vitro, adding to the explicit phenomenological mapping between the SMS and its referent. We assumed SMS are roughly circular, which our observations demonstrated was the case. We first calculated the X and Y extents defined as follows: largest X (east-west or left-right) and Y (north-South or up-down) differences between CELLS at the edge of the SMS, ignoring detached, isolated CELLS. Those two values were averaged to obtain the measure of SMS diameter used to calculate area. This measurement adequately described the trends in SMS growth and remained quite close to the actual area occupied by all CELLS.

2.2.5. Analogue environment on a hexagonal grid

The width of each grid location mapped to 10 μm. Each location was either empty or held a single CELL. A second, identical sized hexagonal grid was overlaid on the first. It contained the NUTRIENT consumed by CELLS, and its value was specified using a floating-point value from 0 to 1. The NUTRIENT within the system mapped primarily to glucose, but other medium components, such as diffusible growth factors, were conflated into the referent. At this early stage, the targeted attributes selected did

not include a requirement for glucose and oxygen being separate components. The single NUTRIENT factor was deemed sufficient. NUTRIENT diffusion used a discretization of the two-dimensional continuous diffusion equation $du/dt = D\nabla^2 u$, where D is the diffusion constant and u is the amount of diffusible material. NUTRIENT moved from high to low density areas. In a given simulation cycle, each location in NUTRIENT space calculated a new value based on the values of itself and its neighbors during the previous cycle using a method adapted for the hexagonal grid from [38]. The new value was $u_{new} = u\,(1 - \lambda) + \lambda\,(u_{NE} + u_{SE} + u_S + u_{SW} + u_{NW} + u_N)/6$, where λ is the discretized diffusion value and u_{NE}, u_{SE}, etc. are the NUTRIENT values at the neighboring locations.

CELLULAR actions, such as NUTRIENT consumption and creating new CELLS (proliferation), occur on larger timescales than diffusion, so for convenience, the NUTRIENT space underwent ten steps for each SEC in CELL space. Having multiple time scales allowed for model accuracy without wasting computation time. Physical and temporal resolutions were purposefully mapped to specific values obtained from the EMT6 system in order to improve simulation realism and help ensure that observed SMS behaviors were not artifacts of unrealistic scaling. The desired time step (Δt) for diffusion within the analogue was related to the unit distance (Δx), D, and λ, such that $\Delta t = 3\Delta x^2\,\lambda/8D$ [39]. For the purposes of the simulation, $\lambda = 0.28$ was chosen in order to allow a Δt of 0.1 SEC when Δx is 10 μm and D is 105 μm^2/SEC [36].

The NUTRIENT space was replenished by an algorithm that detects which empty locations lie outside the SMS and which lie inside. Empty locations outside the SMS were replenished to *initialVal* every ten SEC. We replenished during a simulation more frequently than was done in vitro in order to simulate the effect of stirring within EMT6

cultures: diffusion is not responsible for moving glucose and oxygen toward the EMT6 spheroid during growth, only through it [36]. In order to calculate which depleted spaces resided inside and which were outside, the replenishment algorithm completes multiple passes from the top left corner of the grid to the bottom right and back. On each pass any empty space that is adjacent to an outside empty space is labeled as outside the SMS, sequentially replenishing deeper and deeper fissures. The algorithm used a multiple-pass approach rather than a recursive approach to avoid memory overflow errors. This algorithm was capable of replenishing fissures that extend relatively deep into the SMS.

2.2.6. Analogue is local based

Cells within biological systems evaluate their surroundings through direct interaction with their environment. Most information is transmitted through diffusible signals (which can travel long distances but require contact with a receptor to be recognized), cell environment, or cell-cell interactions. In order to mimic that important biological reality and to preserve a clean separation between mechanism and phenomena we added a key targeted attribute to the list: CELLS must use only local mechanisms (the third attribute in Table 1.1). Each CELL can query the level of NUTRIENT in its local neighborhood along with the characteristics of each neighboring location. In order to locally control SMS surface irregularities and prevent fissure formation the STRESS based movement and proliferation algorithms were developed. They only required CELLS to query their immediate neighbors. Requiring that all mechanisms must be a consequence of local events would falsify some existing individual based models of tumor spheroid growth, such as [39], where an artificial gradient toward the center of the simulated tumor spheroid is created that uses global knowledge.

2.2.7. Initial analogue state

Execution begins with a single CELL that grows unrestricted to 50 CELLS. At that stage, the AXIOMS are implemented. This simplification creates a slight, negligible shift in the growth curves. It should be noted that the initial growth rate of EMT6 spheroids was not clear because EMT6 cells were first grown in a monolayer, and then transferred to cultures dishes until cluster size reached about 400 cells. Only then were they placed in the spinner flasks [21].

2.2.8. CELLS in context

CELLS follow axiomatic operating principles that determine state change, movement, proliferation, and resource consumption. The flow chart in Figure 2.1 demonstrates the full range of actions CELLS can take. CELLS exist in three states: PROLIFERATING, QUIESCENT, and NECROTIC. CELL state is determined by the amount of NUTRIENT to which CELLS are exposed. PROLIFERATING and QUIESCENT CELLS consume NUTRIENT equal to *proConsumeRate* and *quiConsumeRate,* while NECROTIC CELLS do not consume NUTRIENT.

Figure 2.1. Flow chart of action options for an individual CELL.
During each simulation cycle, a CELL begins by checking its state (green diamond). It then checks the conditions of its local environment, as well as values of internal variables to determine intermediate actions (blue diamonds), final actions (yellow diamonds) or CELL death (red diamond). CELLS that move into an empty space will do so deterministically if they are inside the border of the SMS, but stochastically if they are on the outside of the SMS. CELLS that have no neighbors will randomly move outside of the SMS.

Within a simulation cycle, when a CELL finds itself adjacent to an empty space, it will move into that space, simulating random cell movement and churning. That action has the net effect of causing spaces vacated by NECROTIC CELLS to move randomly within the SMS, eventually merging with the external space. The process is illustrated in Figure 2.2. When a CELL is on the outer edge, it will stochastically determine if it moves into an adjacent internal space. The probability of doing so is adjusted based on the CELL'S

STRESS and is biased by the value of *moveEmptyBias*. The higher the STRESS, the greater the likelihood the CELL will move into the empty space, smoothing the local SMS edge. An outside CELL adjacent to an interior space will move into an adjacent space if PRN < p_m. The PRN is drawn from [0,1), and p_m is specified by an empirically derived exponential function of *moveEmptyBias* and STRESS, as detailed below. At constant STRESS, increasing *moveEmptyBias* decreases the likelihood of movement, while if *moveEmptyBias* is constant, increased STRESS will increase the likelihood of movement.

PROLIFERATING CELLS decrement *prolifCounter* during each cycle. When this value drops below zero, the CELL will have an opportunity to create a new CELL. If the CELL is adjacent to empty spaces, it will select one randomly, and then be given an opportunity to place a daughter CELL (a copy) at that location. A CELL given an opportunity to proliferate will do so if PRN < p_b. The PRN is drawn from [0,1), and p_b is specified by an empirically derived exponential function of *proBias* and STRESS, as detailed below. At constant STRESS, the likelihood of proliferation will decrease as *proBias* is increased, and at constant *proBias* increasing STRESS causes qualified CELLS to be less likely to proliferate.

After an attempt at creating a new CELL, *prolifCounter* is reset regardless of whether or not the attempt was successful. A daughter CELL has the same parameter values as the parent, except for *prolifCounter* and *removeCounter*, which are set to unique random values. For simplicity, we specify that CELLS are subject to contact inhibition: only CELLS adjacent to empty space can create new CELLS. Although it is not clear to what extent contact inhibition occurs in vitro, LaRue et al. [28] observed that only the outer two or three cell layers proliferate at the same rate as exponentially growing cells.

Proliferation beyond the SMS surface was not necessary to achieve the targeted attributes (i.e., to survive falsification with the current set of targeted attributes).

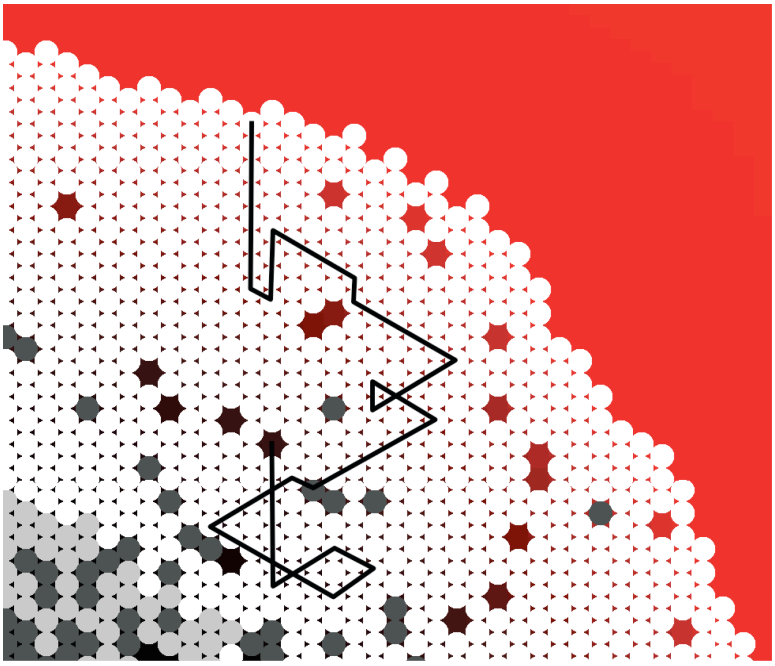


Figure 2.2. Relocation of empty spaces.
Random CELL movement into an adjacent empty space produces the net movement of spaces over time. During each simulation cycle, any CELL adjacent to an internal empty space move into that space. Consequently, empty spaces move randomly (black line) within the SMS until they exit at the outside edge. Spaces may be moved multiple times during a single simulation cycle.

2.2.9. Stochastic cell movement and proliferation

When a CELL on the outside of the SMS is adjacent to an empty space it will move into that space if a PRN < p_m, where PRN is a pseudo random number in the interval [0,1). The value of p_m is determined by the function $p_m = \left(\frac{S+3}{6}\right)^{7m+1}$, where S is the STRESS experienced by the CELL and m is the value of *moveEmptyBias*. This function has the effect of skewing the distribution of p_m based on *moveEmptyBias*. The higher the value of *moveEmptyBias* the more extreme the skewing will be, so that CELLS with a low STRESS and a high *moveEmptyBias* will almost never move into adjacent empty spaces, while CELLS with a high STRESS will do so regardless of the value of *moveEmptyBias*.

CELLULAR proliferation is biased in a similar fashion, except that in this case CELLS with a high STRESS will be less likely to create new CELLS. CELLS will proliferate if PRN < p_b, with PRN being between [0,1). The equation used to determine the value of p_b is $p_b = \left(\frac{6-S}{8}\right)^{7b+1}$, with S being the STRESS of the CELL and b being the value of *proBias*. The skewing effect is similar, but the influence of STRESS is reversed, so CELLS with a high STRESS and high *proBias* will almost never create new CELLS, while CELLS with low STRESS will be very likely to create new CELLS regardless of the value of *proBias*.

2.2.10. METABOLISM requires a single source of NUTRIENT

An SMS differs significantly from the individual based models of Chignola et al. [10] and Schaller et al. [6], especially in its simple representation of metabolism. The main similarities are that SMS use a diffusible NUTRIENT and CELLS DIE when insufficient NUTRIENT is available. There was no need to represent a particular type of metabolism (aerobic or anaerobic), only that CELLS consume NUTRIENT equal to *proConsume* or *quiConsume* and change state based on NUTRIENT level. We achieved the targeted attributes without being forced to add additional METABOLIC complexity. Because we achieved those attributes using a simple representation, we can achieve the same behaviors using a more complicated representation of metabolism.

2.2.11. Fissure formation is related to STRESS

Early SMS versions had no means to control shape, either directly or indirectly, leading to fissure formation similar to that seen in [40]. SMS fissures were induced by diffusion-limited aggregation (DLA). DLA is a phenomenon that occurs when objects

move randomly in space until they encounter and adhere to each other, forming structures with crystalline appearance [41]. SMS fissures form as the empty spaces move about because of CELL movement, and adhere when another space is encountered. Empty spaces, even though they are not actively moving objects, are subject to DLA rules because they effectively walk randomly through the SMS. Fissures form as spaces connect to each other. The inner extreme of a fissure is the closest outside spaces to the NECROTIC core, where spaces are generated. In order to prevent fissure formation, we developed the STRESS-based proliferation algorithm. It helps prevent fissure development. CELLS at the inner extreme of any fissure will have low STRESS values, leading to preferential proliferation at that location.

In order to avoid extreme, abiotic SMS surface irregularities, each CELL creates new CELLS and moves based on its STRESS value. STRESS maps somewhat to the adhesion mechanisms used in [2, 6]. It also maps to a combination of surface tension and adhesion. However, it was only necessary to have it operate at the SMS surface. CELLS having smaller STRESS values will be more likely to create new CELLS and less likely to move inward into empty spaces. To calculate STRESS, a CELL uses a two-pass algorithm in each cycle. First, it determines its *initialStress*: it subtracts two from the number of empty spaces in the neighborhood. During the second pass, a CELL counts the number of CELLS in its neighborhood that are on the SMS edge (*outsideNeighbors*). As illustrated in Figure 2.3, CELLS then calculate their final stress depending on *outsideNeighbors*. If $outsideNeighbors \neq 2$, $finalStress = initialStress + 1$. If $outsideNeighbors = 2$, the CELL queries these two neighbors and sums their *initialStress*, and if that sum < 0, $finalStress = initialStress - 1$. If the sum is > 1, $finalStress = initialStress + 1$, and if the sum is 0 or 1,

finalStress = *initialStress*. This algorithm has the effect of transmitting the STRESS felt by one CELL to its neighbors, enabling CELLS to have different final STRESS values even if their neighborhoods are identical.

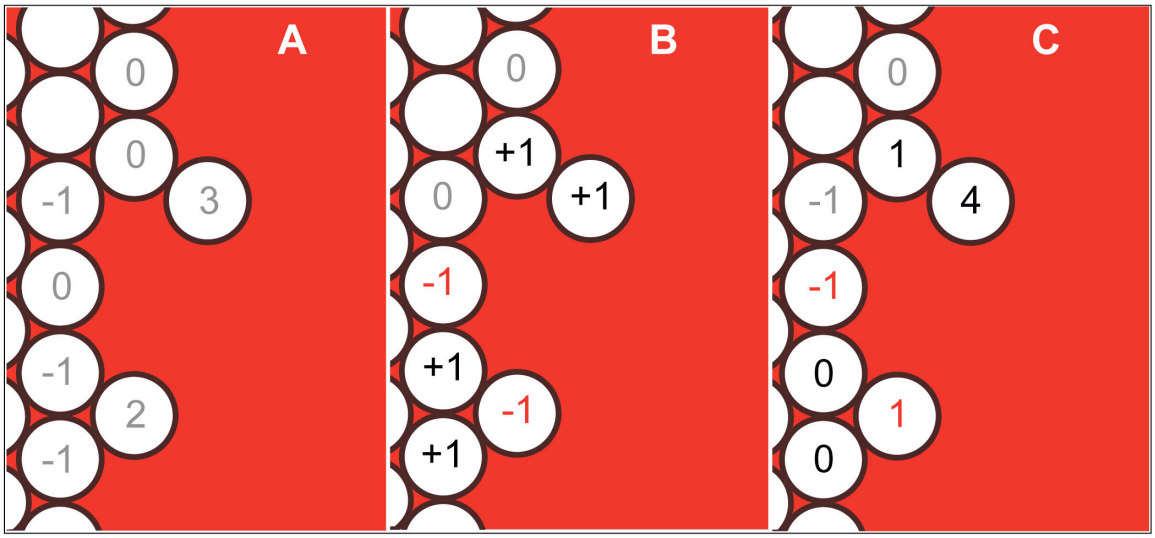


Figure 2.3. Illustration of a CELL determining its level of STRESS.
(A) *InitialStress* is calculated based on the number of empty spaces. (B) The change in STRESS is calculated based on number of outside neighbors and their *initialStress* values, with some CELLS increasing in STRESS (black values), some decreasing (red values) and others staying the same (gray values). (C) STRESS is calculated by summing the value of *initialStress* and the change in the value of STRESS.

Figure 2.4 shows sequential screen shots of the stress felt by CELLS in a growing SMS and demonstrates the preferential nature of proliferation. Figure 2.4A shows the initial arrangement. In Figure 2.4B, the starred CELL has moved to fill an empty space, changing the local CELL arrangement and each CELL'S resulting STRESS value. Consequently, the starred CELL has a very low STRESS value and a corresponding higher chance of proliferating. When its *prolifCounter* reaches zero, the starred CELL creates a new CELL, as illustrated in Figure 2.4C, returning that portion of the SMS to its original arrangement.

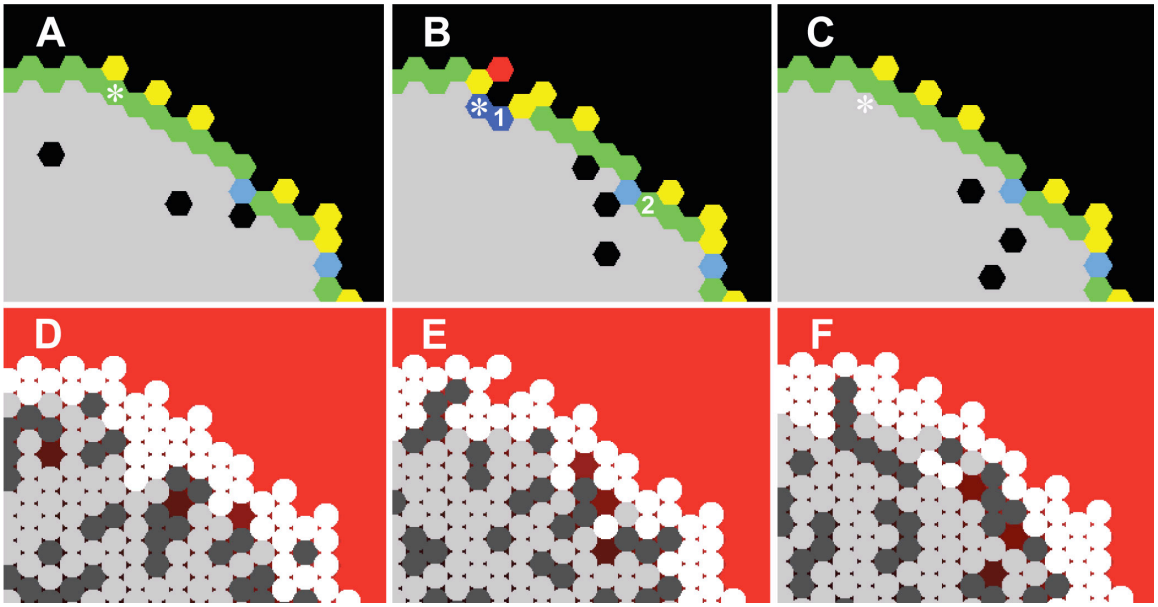


Figure 2.4. Illustration of CELLS responding to STRESS at low NUTRIENT.
(A-C) Illustrations of STRESS levels at sequential time steps. Only CELLS at the surface are color-coded. STRESS levels: dark blue = –2, light blue = –1, green = 0, yellow = 1, orange = 2, and red = 3. (A) During one simulation cycle, the empty space below and to the left of the starred (*) CELL is adjacent to that CELL; the CELL then moves inward to fill that empty space. (B) During the next simulation cycle, the starred CELL has a low STRESS and so becomes likely to create a new CELL. The stress algorithm allows CELLS that have equivalent immediate neighborhoods, such as the CELLS labeled 1 and 2, to have different STRESS values. Because the neighbors of CELL 1 have higher *initialStress* values than the neighbors of CELL 2, CELL 1 will have a higher STRESS and be more likely to create a new CELL during the simulation. (C) During the third simulation cycle the starred CELL creates a new CELL, places it in the adjacent space, resulting in a return to initial conditions. (D-F): CELL state view at equivalent TIME steps.

2.2.12. Shedding of cells from the SMS surface

An occasional CELL will become isolated near the SMS surface because of normal AXIOM operation. In order to prevent their local accumulation, we implemented an algorithm that simulates shedding and the consequences of shear force caused by stirring. Any CELL that has no CELL neighbors will move randomly, selecting one of its six immediate neighbors using a uniform distribution, stopping when it encounters another CELL. Most CELLS reattach to the SMS or form small clusters. An occasional CELL will move far enough to exit the grid; it is then removed from the simulation. The number of CELLS shed currently is significantly smaller than that reported in [35]. Shedding was not deemed a sufficiently important attribute to target at this stage. Should it be targeted, it

will be straightforward to add a shedding AXIOM and then adjust parameter values to reestablish matching behaviors. As an example, a similar modification was carried out between an earlier version of the analogue [42] and the current one. Cells in [42] consumed the same quantity of NUTRIENT regardless if they were in the QUIESCENT or PROLIFERATING state, but in the current analogue the amount consumed is different.

2.3. Results

CELL growth leads to formation of Simulated MULTICELLULAR Spheroids (SMS). Measurements of SMS attributes during execution mimic characteristics of EMT6 spheroid growth. Figure 2.5 shows an example qualitative measurement of the SMS as a two dimensional cross-section. It shows that SMS displayed the characteristic layered structure of EMT6 spheroids. The PROLIFERATING layer appears thicker than is often seen in EMT6 spheroids, but note that while the majority of CELLS in the VIABLE rim are in the PROLIFERATING state, only the CELLS on the outer layer of the SMS are actively creating new CELLS. Figure 2.6 shows that SMS growth curves were similar to reported EMT6 spheroid growth curves when CELLS used the parameters listed in Table 2.1 and the nine axiomatic operating principles listed in Table 2.2. AXIOM application was rigid in the sense that when a precondition was met, the appropriate AXIOM was always applied. AXIOMS 7 and 9 were stochastic. During a simulation cycle, a CELL could subsequently apply more than one AXIOM, such as 1, 4, 7, and 9 for a PROLIFERATING CELL or 3 and 5 for a NECROTIC CELL.

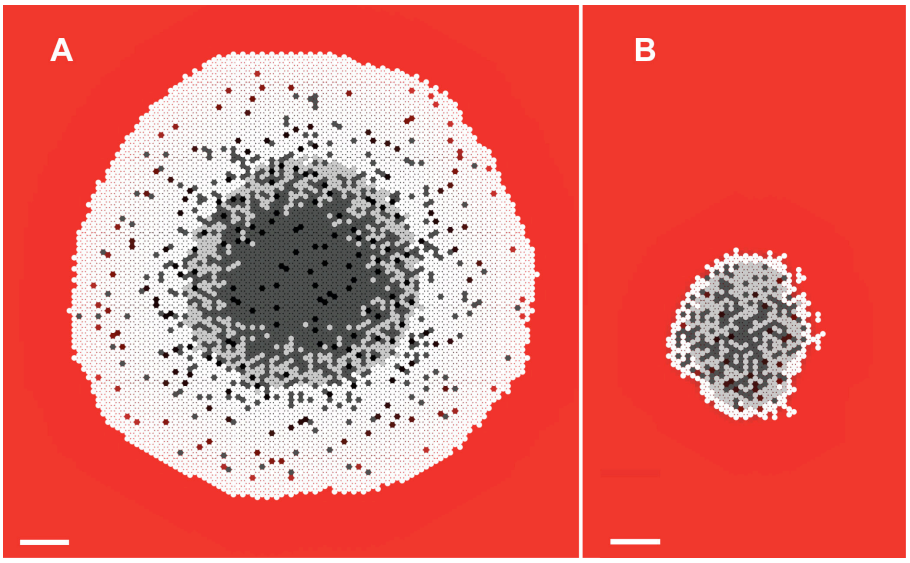


Figure 2.5. SMS cross-sections at 17 DAYS.
Scale bar: 100 μm. Parameter values were those listed in Table 2.1. White circles: proliferating CELLS; light gray circles: quiescent CELLS; dark gray circles: NECROTIC CELLS. The background gradient (from red to black) represents NUTRIENT levels relative to the maximum value in red. (A) Growth occurred at high NUTRIENT, which maps to 0.28 mM oxygen and 16.5 mM glucose. (B) Growth occurred at low NUTRIENT, which maps to 0.08 mM oxygen and 0.8 mM glucose.

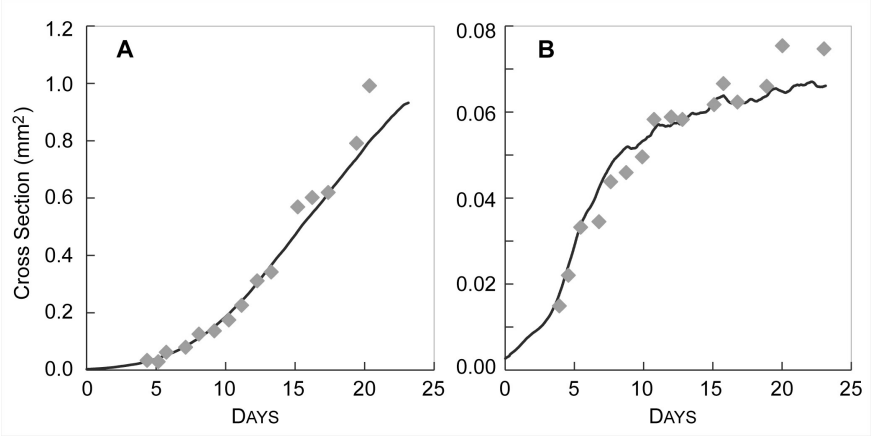


Figure 2.6. EMT6 and SMS growth curves.
In vitro growth values (gray diamonds) were adapted from [21] by calculating spheroid diameters from measured volumes, assuming a circular cross-section. SMS values were obtained by specifying that CELL diameter maps to 10 μm, measuring the greatest X, Y extents, excluding isolated CELLS, and assuming a circular cross-section. Parameter values were those listed in Table 2.1. (A) SMS growth at high NUTRIENT. Values are means of ten runs. EMT6 spheroid values were from [21] at 0.28 mM oxygen and 16.5 mM glucose. (B) SMS growth was under low NUTRIENT. Values are means of ten runs. EMT6 spheroid values from [21] at 0.07 mM oxygen and 0.8 mM glucose.

2.3.1. In silico growth curves matched in vitro growth curves

For the parameter values listed in Table 2.1, SMS growth curves were quantitatively similar to those of EMT6/Ro spheroids for both high and low nutrient

conditions. CELLS within an SMS proliferated initially at an exponential pace. Growth then slowed and became linear because only CELLS near the outer SMS rim could reproduce. The increase in cross-sectional area was linear until CELLULAR NECROSIS began. Thereafter, SMS growth rate began decreasing toward zero. A stable size was reached when CELL creation was balanced by CELL removal. Plots of SMS cross-sectional area over time (Figure 2.6) closely mirrored EMT6 spheroid growth [21] for both high and low levels of NUTRIENT. For simplicity, as discussed under Methods, we conflated measured concentrations of glucose and oxygen, along with the other in vitro nutrients, and represented the entire collection using the factor NUTRIENT. High NUTRIENT level mapped to 16.5 mM glucose and 0.28 mM oxygen. Low NUTRIENT level mapped to 0.8 mM glucose and 0.07 mM oxygen. The reported coefficient of variation of mean cross-sectional EMT6 spheroid area between multiple in vitro experiments was roughly 29% at 7 days, increasing over time [28]. Given that, and the fact that EMT6 spheroids increase their size by many orders of magnitude during growth, we judged that having simulated values within 15% of referent values would be reasonable, and made that a targeted attribute (Table 1.1). The mean percent error between in silico and in vitro data was 12% for high and 8% for low NUTRIENT, which was within the targeted 15% range. The only parameter changed between the two conditions was *initialVal*, the level of NUTRIENT present outside the SMS during the simulation. For both conditions, CELLS used the same Table 2.2 set of operating principles. Under Methods, we describe that only minimal tuning of the indicated subset of the parameters in Table 2.1 was needed to achieve these matching growth characteristics.

2.3.2. In silico doubling times were similar to in vitro doubling times

The measured doubling times for SMS and EMT6 spheroids were similar at high NUTRIENT levels, but quite different at low NUTRIENT levels, as shown in Table 2.3. At low nutrient levels, EMT6 cell number doubled every 17 hours, whereas the SMS required 40 HOURS. This apparent discrepancy was initially difficult to explain, considering that the growth curves were very similar. Some explanatory factors may include the high variability of the in silico results at small SMS sizes, as well as the experimental variability between in vitro trials. In addition, initial doubling times were calculated [21] using a best fit of the Gompertz equation [43] to data from multiple experiments. The Gompertz equation describes an exponential curve with an exponentially decreasing growth rate. It can be fit to many types of in vitro animal and tissue growth data. The equation used was $y = ae^{-e^{(b-ct)}}$, where t is time, y is cross-sectional area, and a, b, and c, are the parameters of the equation. When we fit the in vitro results (from the single experiment we used for validation) to the Gompertz equation, we obtained an initial doubling time of 21.6 hours at high nutrient levels and 26.4 hours at low, as compared to 19.2 HOURS for in silico simulations at high and low NUTRIENT, as shown in Table 2.3. These results are more consistent, which is not surprising, as we tuned *prolifDelay* and *proBias* to generate in silico doubling times that could be mapped approximately 1:1 to wet-lab doubling times. *ProlifDelay*, a stochastic parameter, is the average TIME interval that a CELL must wait before it has the option to create a new CELL. The value of *prolifCounter* specifies the interval for each CELL. Each *prolifCounter* value is calculated from *prolifDelay* as described under Methods. To enable a successful proliferation event, a pseudo-random number must exceed a specified

value. The variable *proBias* specifies the probability distribution from which that value is drawn, as described under Methods. The larger the *proBias* value, the more extreme that distribution, making it increasingly difficult for a new CELL to be created.

Table 2.3. Comparison of in vitro and in silico growth characteristics.

Condition	Initial doubling time	Viable rim width	Necrosis onset time	Necrosis onset size	Maximum area
In silico					
High NUTRIENT*	21.4 HOURS / 19.2 HOURS‡	245 μm	11.4 DAYS	0.253 mm^2	1.61 mm^2 ‡ / 1.46 mm^2‡
Low NUTRIENT*	40.0 HOURS / 19.2 HOURS‡	62 μm	4.4 DAYS	0.0266 mm^2	0.0645 mm^2‡
In vitro					
16.5 mM glucose & 0.28 mM oxygen	23.0 hours† / 21.6 hours‡	240 μm	N/A	0.209 mm^2	3.25 mm^2 † / 2.79 mm^2‡
0.8 mM glucose & 0.07 mM oxygen	17.0 hours† / 26.4 hours‡	60 μm	N/A	0.0181 mm^2	0.221 mm^2 †/ 0.0725 mm^2‡

*: Values for mean of ten runs. †: From Gompertz fit by Freyer and Sutherland. ‡: From Gompertz fit by authors.

2.3.3. Measured viable rim widths were similar

Viable cell rim widths, eighth in Table 1.1, have been characterized, and were used to further validate SMS attributes. The data in Table 2.3 show that VIABLE SMS rim widths were close to in vitro values. Because the coefficient of variation of EMT6 spheroid areas was at least 29%, the corresponding value for radius was roughly 15%. We specified that any SMS radius within 15% of a referent radius would be acceptably similar because that radius would be experimentally indistinguishable from a repeat EMT6 experiment, had one been preformed. We observed a mean SMS rim width under the high NUTRIENT condition that mapped to 245 μm, compared to 240 μm in vitro, a difference of 2%. At low NUTRIENT, mean SMS rim width mapped to 62 μm, compared to 60 μm in vitro, a 3.3% difference. Although we performed some tuning of the critical levels required to remain in the PROLIFERATING or QUIESCENT state, these similarities are still noteworthy. They reinforce the likelihood that the principles of operation used by SMS CELLS may map to a corresponding set of operating principles used by EMT6 cells.

2.3.4. NECROSIS onset and final saturation size were similar

Under high NUTRIENT, measures of SMS diameters at NECROSIS onset, listed in Table 2.3, achieved the targeted similarity measure; they were within 15% of those observed by Freyer and Sutherland [21]. Following SMS execution under high NUTRIENT, mean diameter at which the SMS first underwent NECROSIS mapped to roughly 530 μm, compared to 516 μm for the EMT6 spheroids under comparable conditions, a 2.7% difference. Under the low NUTRIENT condition, SMS underwent NECROSIS when the system reached a diameter corresponding to approximately 180 μm, compared to 152 μm for EMT6 spheroids, an 18.4% difference.

The maximum sizes attained by SMS were not similar to those predicted by Freyer and Sutherland [21], (Table 2.3). Freyer and Sutherland did not measure maximum sizes, but instead inferred them by fitting data to the Gompertz equation and then using the fitted equation to predict an expected maximum size. The data fit were averages of results from experiments on different batches of EMT6 spheroids. Because the SMS were being compared to data from a single experiment, we fit the Gompertz equation to that referent data (Figure 2.6). Table 2.3 shows that the new result did not differ significantly from the one originally reported at high nutrient levels: the moderate in silico–in vitro discrepancy remained. However, the maximum fit size was smaller at low nutrient concentrations, down from 0.221 mm^2 to 0.0725 mm^2. The maximum size reached by the SMS at low nutrient mapped to 0.0645 mm^2, a difference of 11%. That was judged acceptably similar to our Gompertz equation fit. A reasonable conjecture for the discrepancy at high nutrient levels is that the set of operating principles used by cells in maturing EMT6 spheroids were somewhat different than the set used earlier, during spheroid expansion. Our goal was to seek one set of SMS operating principles that would

enable validation for both high and low NUTRIENT conditions. It would be straightforward to relax that requirement and achieve improved similarity at high NUTRIENT levels.

2.3.5. SMS shape and stability were controlled by *proBias*

How important are the quantitative aspects of the AXIOMS in controlling spheroid shape and stability? AXIOMS 7 and 9 in Table 2.2 play critical roles in controlling SMS shape and stability. The consequences of their application in concert with the other seven depend to a large degree on the value assigned to the parameter *proBias*, which, as explained under Methods, influences the likelihood of stressed CELLS to proliferate. We conducted experiments at varied levels of *proBias*. Results are shown in Figure 2.7. Increasing *proBias* improved an SMS's ability to fill in fissures that formed after growth stabilization. Low levels of *proBias* enabled fissures to reach toward the SMS center, destabilizing the structure and causing chaotic, uncontrolled growth. We have found no evidence that this destabilizing mechanism maps to in vitro counterparts. However, it may demonstrate a principle: surface irregularities can affect a spheroid's growth rate. An SMS that has elongated to form a rod-like structure could not easily increase its width because of limited NUTRIENT availability. However, absent other constraints, nothing would prevent it from elongating further. Clearly, there are other factors and forces involved in maintaining the stability and shape of large spheroids in vitro. However, they are beyond the SMS's current scope. SMS are capable of maintaining stable shapes for a specified TIME within certain parameter ranges, outside of which new AXIOMS and/or non-local constraints would be required.

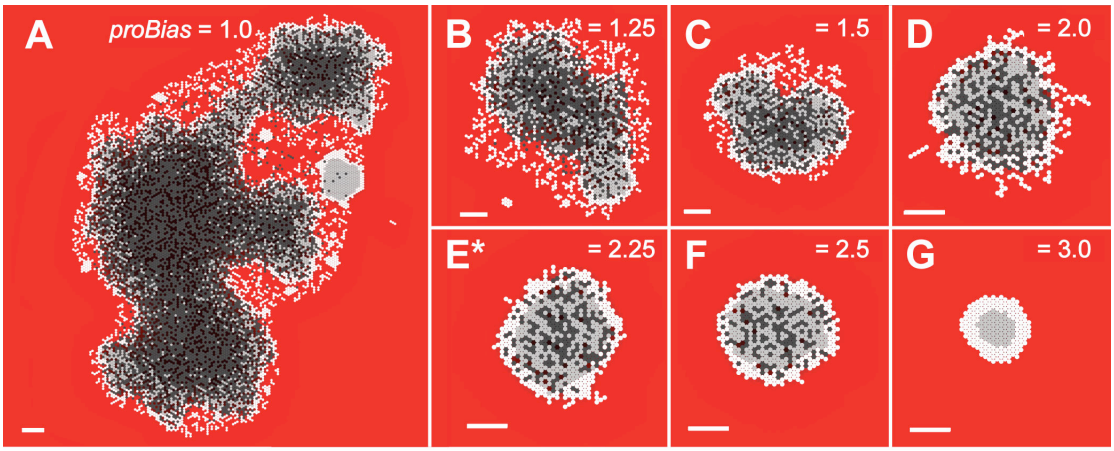


Figure 2.7. SMS cross-sections at varied *proBias* values and low NUTRIENT.
All images were recorded at 18 DAYS. Scale bar: 100 μm. Other parameter values were as listed in Table 2.1. (A)–(G) *proBias* values are shown. *: *proBias* value in Table 2.1.

2.3.6. SMS long-term shape changes lead to instability

The AXIOMS used to manage CELL STRESS were effective at maintaining stable small and medium sized SMS. However, when growth was extended beyond 50 DAYS under high NUTRIENT, SMS began to lose their circular shapes. The 67-DAY old SMS in Figure 2.8 is an example. Because of the stochastic nature of the events involving each CELL, the growth trajectories, shapes, and sizes of separate SMS executions can be different under identical conditions. For the same reason, small regional differences in large, mature SMS can gradually become amplified, resulting in large subregions having measurably different characteristics. This leads to an unstable system. Measures of SMS long-term growth at high and low NUTRIENT are provided in Figure 2.9. EMT6 spheroids do exhibit varied shapes during growth. However, we are not aware of in vitro observations that can be used to validate this SMS behavior, possibly because it is challenging (and expensive) to maintain large EMT6 spheroids in culture for 50 days or more (although Chignola et al. have maintained Rat 9L spheroids to 70 days [30]). If maintenance of generally circular shapes beyond 50 days under high NUTRIENT were to be added to the targeted attributes list, the current SMS would be falsified. Inclusion of additional mechanisms would be needed to reestablish validation.

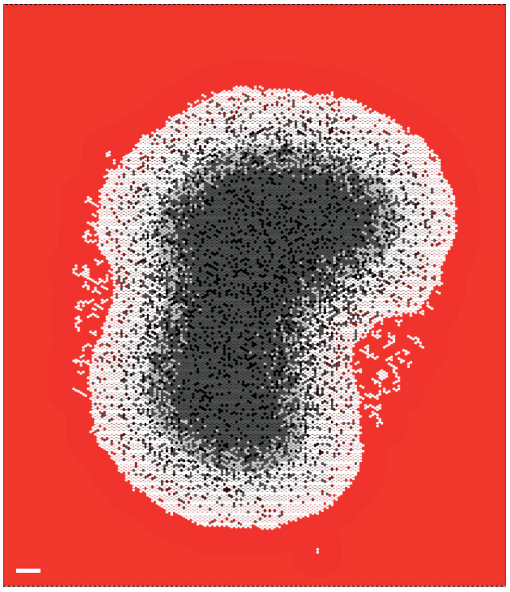


Figure 2.8. An SMS cross-section at 67 DAYS at high NUTRIENT level.
SMS shape is no longer circular. Scale bar 100 µm. Parameter values were those listed in Table 2.1.

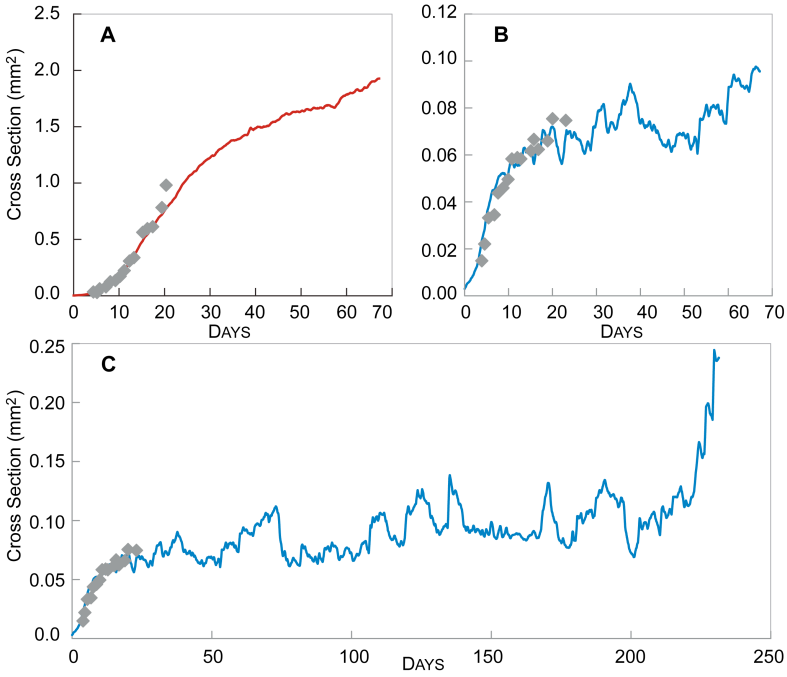


Figure 2.9. Results from three extended SMS experiments.
Gray circles: in vitro data as in Figure 2.6. Parameter values were those listed in Table 2.1. (A) SMS growth at high NUTRIENT for 67 DAYS. SMS growth at low NUTRIENT for 67 (B) and 230 (C) DAYS. Eventual maximum size before destabilization for both (A) and (B) is similar to that predicted by fitting the growth of multiple runs to the Gompertz equation, as observed in Table 2.3. In (C) the SMS remains stable in size for an extended period of time, but eventually destabilizes.

2.3.7. Varying parameters changed growth curve and SMS shape

We conducted experiments in which we varied parameter values and observed the effect on measures of SMS growth and morphology. The results, summarized in Table 2.4, indicate whether increasing a parameter increased, decreased, or did not affect a specific measure (such as maximum size reached). In addition, we examined the consequences of changing parameter values in more detail. Changing *moveEmptyBias* had a limited but significant effect on SMS morphology (Figure 2.10) and growth (Figure 2.11). The parameter *moveEmptyBias* influenced movement of CELLS exposed to the outside surface of the SMS or adjacent to a fissure. The larger the value of *moveEmptyBias*, the less likely a CELL at the edge experiencing low STRESS (defined under Methods) would move into an adjacent empty space when given the opportunity. A larger *moveEmptyBias* value hindered fissure elongation. *MoveEmptyBias* was tuned empirically to control SMS shape but still allow CELL-free spaces to exit the SMS rather than be trapped inside for an extended duration. CELLS with larger *moveEmptyBias* values experiencing low STRESS rarely moved into adjacent empty spaces, whereas CELLS under high STRESS are likely to do so.

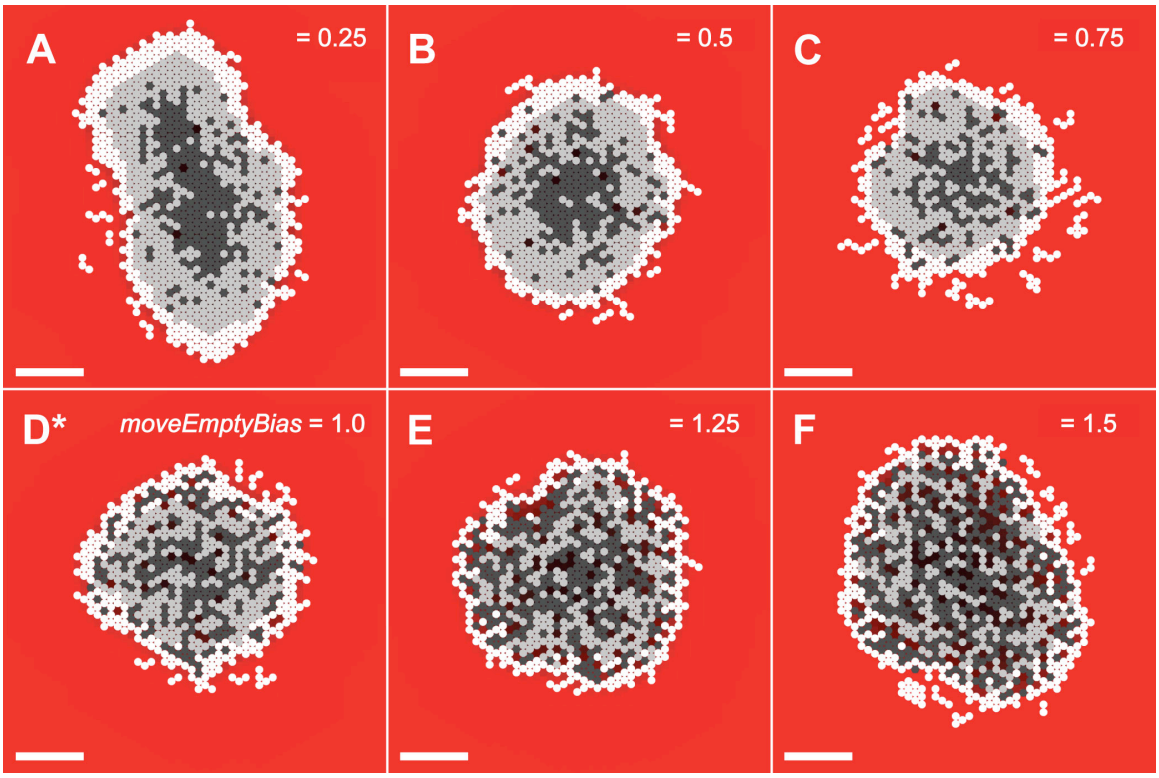


Figure 2.10. SMS cross-sections at varied *moveEmptyBias* values and low NUTRIENT.
All images were recorded at 21 DAYS. Scale bar: 100 µm. Other parameter values were as listed in Table 2.1. (A)–(F) *moveEmptyBias* values are shown. *: *moveEmptyBias* value in Table 2.1. Cross-sections at *moveEmptyBias* = 0 are not shown because they grew too quickly and filled the available space before 21 DAYS elapsed. As *moveEmptyBias* increased, more empty spaces were visible within the SMS.

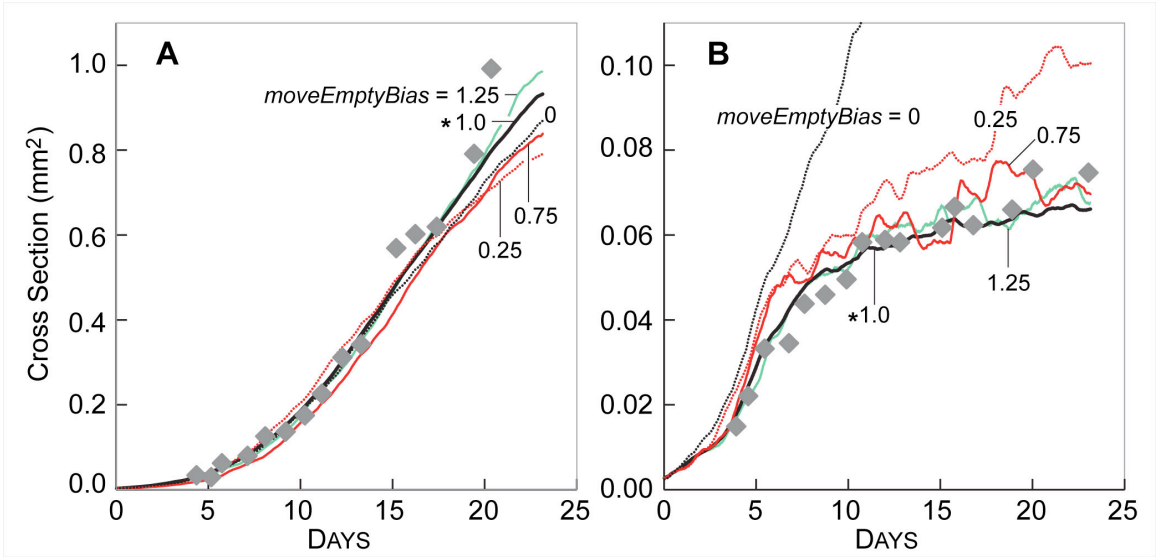


Figure 2.11. Influence of *moveEmptyBias* on SMS growth.
Gray diamonds: in vitro data as in Figure 2.6. Other parameter values were those listed in Table 2.1. Colored lines are results of single experiments for the indicated values of *moveEmptyBias* from 0 to 1.5 (*moveEmptyBias* = 0 plus the same values as in Figure 2.10). (A) high NUTRIENT; (B) low NUTRIENT. *: *moveEmptyBias* value in Table 2.1.

As demonstrated by Figure 2.11, when *moveEmptyBias* was set to zero, SMS grew linearly at a high rate in low NUTRIENT conditions and failed to saturate. Fissures appeared, which caused the SMS to destabilize and grow chaotically. Increasing *moveEmptyBias* by as little as 0.25 resulted in almost complete SMS saturation. Further increasing *moveEmptyBias* did not significantly affect growth rates or stability, but small changes were evident at both low and high NUTRIENT levels. While *moveEmptyBias* does not directly map to an in vitro quantity, these results indicate that there may be threshold values for shape maintenance mechanisms, below which an EMT6 spheroid would generally become unstable.

Table 2.4. Effects of increasing parameters on in silico measures.

Parameter	Maximum size	Viable rim width	Quiescence onset	Necrosis onset	Growth rate	Doubling time	Stability	Necrotic core size
diffusionRate	↑	↑	↑	↑	—	—	↔	↔
initialVal	↑	↑	↑	↑	↓	↓	↔	↑
proConsumeRate	↓	↓	↓	↓	—	—	—	↔
quiConsumeRate	↓	↓	—	↔	—	—	—	↔
prolifDelay	↔	—	↑	↑	↓	↓	—	↔
removeDelay	↑	—	—	—	—	—	↔	↑
proNut	↔ / —	—	↓	↓	—	—	↓	↔
quiNut	↓	↓	—	↓	↓	—	—	— / ↓
proBias	↓	↑	↑	↑	↓	↓	— / ↑	↓
moveEmptyBias	↔	↓ / ↑	—	—	↑	—	↑	↑

↑: Positive effect. ↓: Negative effect. ↔: Effect limited or not easily determined. —: No effect. Cells with two values (e.g. — / ↑) indicate different effects at high nutrient (left) and low nutrient (right). Cells with one value indicate similar effect at high and low nutrient. Quiescence onset and necrosis onset are the measured times of onset. A positive effect indicates that the time of quiescence or necrosis onset increased, thus they occurred later in the simulation.

The value of *quiConsumeRate* determined the amount of NUTRIENT per SECOND consumed by each quiescent CELL. The value was tuned between zero and *proConsumeRate*. Varying *quiConsumeRate* produced consistent and dramatic results. Increasing the amount of NUTRIENT consumed by QUIESCENT CELLS reduced the number of QUIESCENT CELLS capable of existing within the SMS. The consequences are clearly

visible in Figure 2.12: as *quiConsumeRate* increased, the width of the viable rim decreased, as did the number of QUIESCENT CELLS in the system and overall SMS size. Figure 2.13 shows that the growth rate and saturation size steadily decreased. *ProConsumeRate* specified the amount of NUTRIENT per SECOND consumed by proliferating CELLS. For simplicity, PROLIFERATING CELLS consumed the same amount of NUTRIENT regardless of whether they were actively creating a new CELL, waiting for an opportunity to do so, or were unable to do so because other CELLS surrounded them. *ProConsumeRate's* value in Table 2.1 was purposefully selected to be within the range of glucose consumption rates reported in [36].

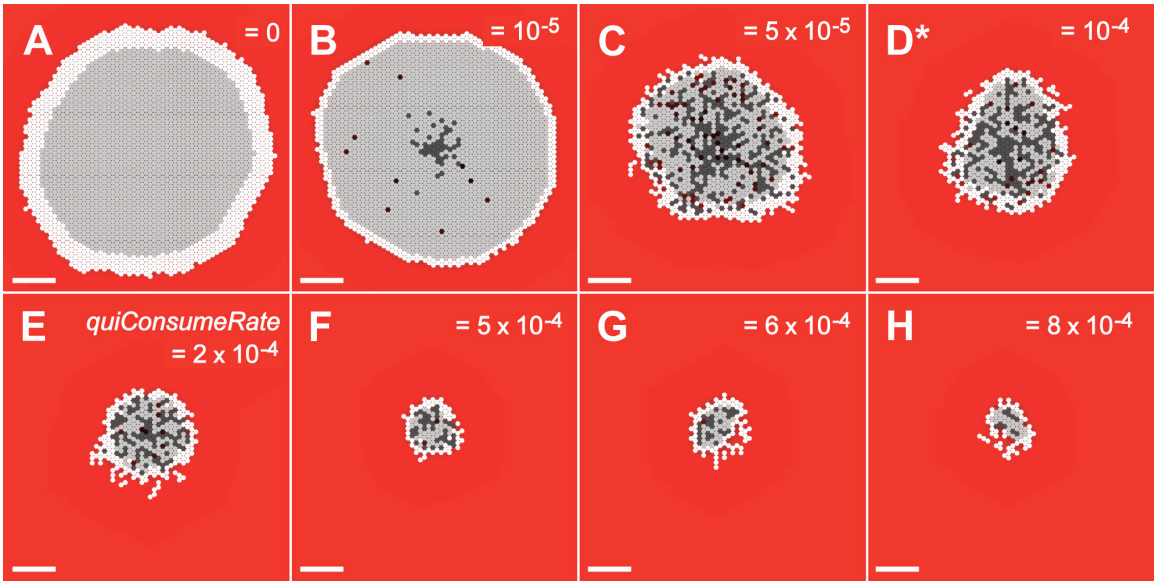


Figure 2.12. SMS cross-sections at varied *quiConsumeRate* and low NUTRIENT.
All images were recorded at 13 DAYS. Scale bar: 100 μm. Except for *quiConsumeRate*, parameter values were those listed in Table 2.1. (A)–(H) *quiConsumeRate* values are shown. *: *quiConsumeRate* value in Table 2.1.

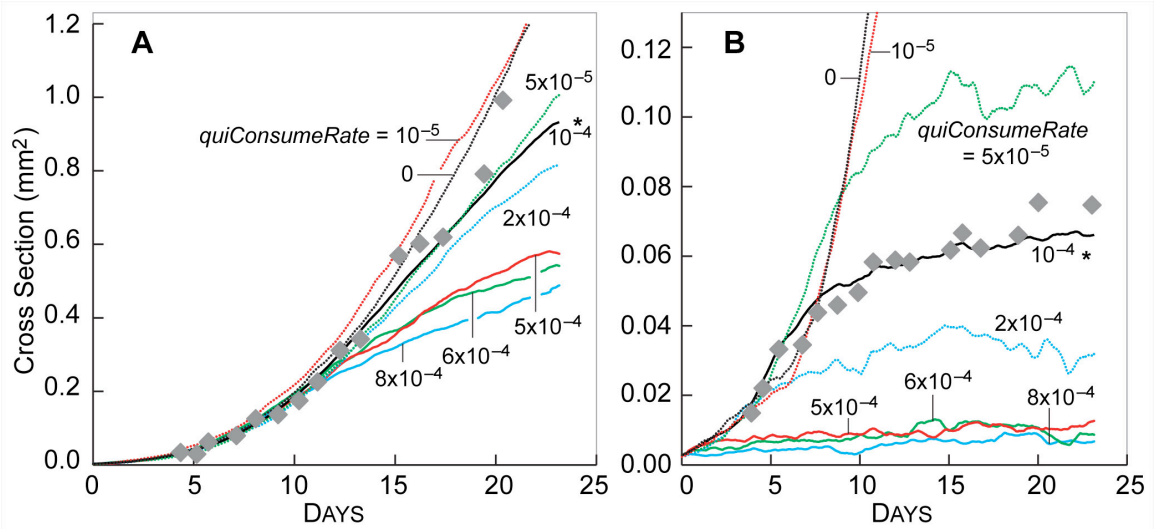


Figure 2.13. Influence of *quiConsumeRate* on SMS growth.
Gray diamonds: in vitro data as in Figure 2.6. Other parameter values were those listed in Table 2.1. Colored lines are results of single experiments for the indicated values of *quiConsumeRate* from 0 to 8.0 x 10^{-4} (same values as in Figure 2.12). (A) high NUTRIENT; (B) low NUTRIENT. *: *quiConsumeRate* value in Table 2.1.

A CELL switched from PROLIFERATING to QUIESCENT state when the amount of NUTRIENT at its location dropped below the value of the parameter *proNut*. If NUTRIENT later increased above *proNut*, the CELL returned to the PROLIFERATING state. Changing the value of *proNut* changed the amount of NUTRIENT that CELLS required to remain in the PROLIFERATING state. When set to 8.0 x 10^{-4}, CELLS transitioned directly from the PROLIFERATING to the NECROTIC state, as shown in Figure 2.14. We do not distinguish between simulated necrotic and apoptotic cell death, instead conflating both into removal of NECROTIC CELLS. When referring to in vitro research we defer to the original documents for terminology. CELL growth at that setting under low NUTRIENT (Figure 2.15) was low, as PROLIFERATING CELLS consume more NUTRIENT than QUIESCENT CELLS. Increasing *proNut* to 2.0 x 10^{-3} produced little change in morphology or growth rate, but both measures changed dramatically when *proNut* was raised to 3.0 x 10^{-3}. At that value, a population of QUIESCENT cells became clearly evident, and the growth rate and stable maximum size was noticeably larger. That trend did not continue, however. As *proNut*

increased further, first to 4.0 x 10^{-3} and then to 5.0 x 10^{-3}, only small changes in morphology and growth curves were evident. The population of QUIESCENT CELLS was only slightly larger. Another sharp change was evident as *proNut* reached 6.0 x 10^{-3}: the SMS destabilized (Figure 2.14G) and the growth curve did not plateau (Figure 2.15B). The results suggested that a window existed within which the number of PROLIFERATING CELLS, having higher consumption rates, balanced the number of QUIESCENT CELLS, which had lower consumption rates. When the level of NUTRIENT within a location dropped below the value *quiNut*, the CELL switched irreversibly to the NECROTIC state. As is evident from Figure 2.16 and Figure 2.17, varying *quiNut* had less complex effects. As *quiNut* was increased, the growth rate, saturation size, and viable rim width all decreased.

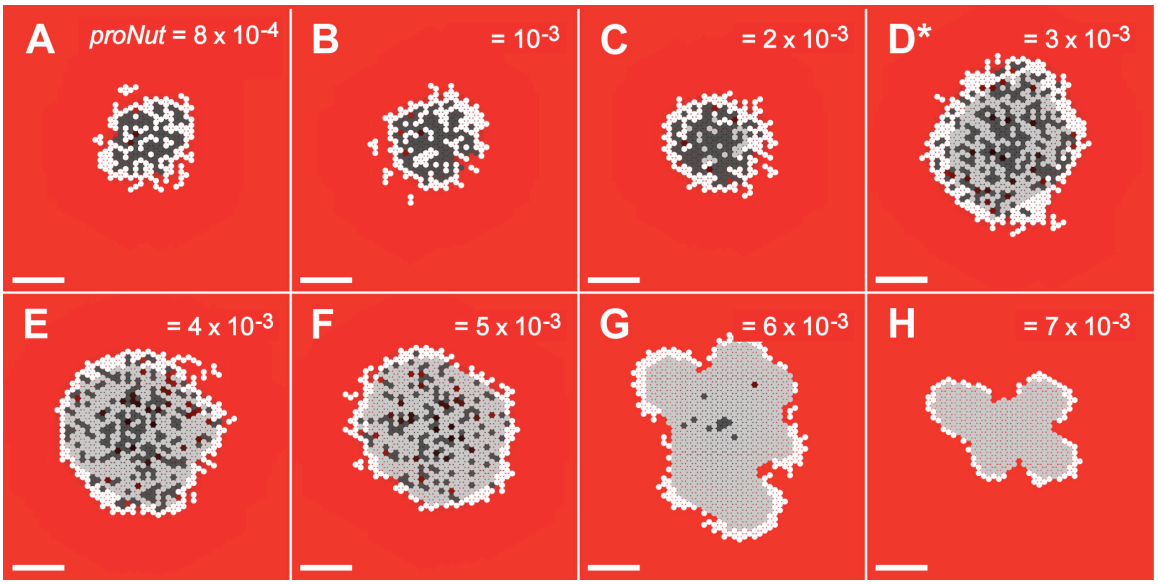


Figure 2.14. SMS cross-sections at varied *proNut* and low NUTRIENT.
All images were recorded at 18 DAYS. Scale bar: 100 µm. Other parameter values were those listed in Table 2.1. (A)–(H) *proNut* values are shown. *: *proNut* value in Table 2.1. Note that while size increased initially with increasing *proNut* (a consequence of increased numbers of lower-consumption rate quiescent CELLS), larger *proNut* values caused the proliferating rim to become so thin that the SMS destabilized. The more quickly CELLS that are incapable of proliferating transition to the (lower-consumption) QUIESCENT state, the larger the size of the stable SMS.

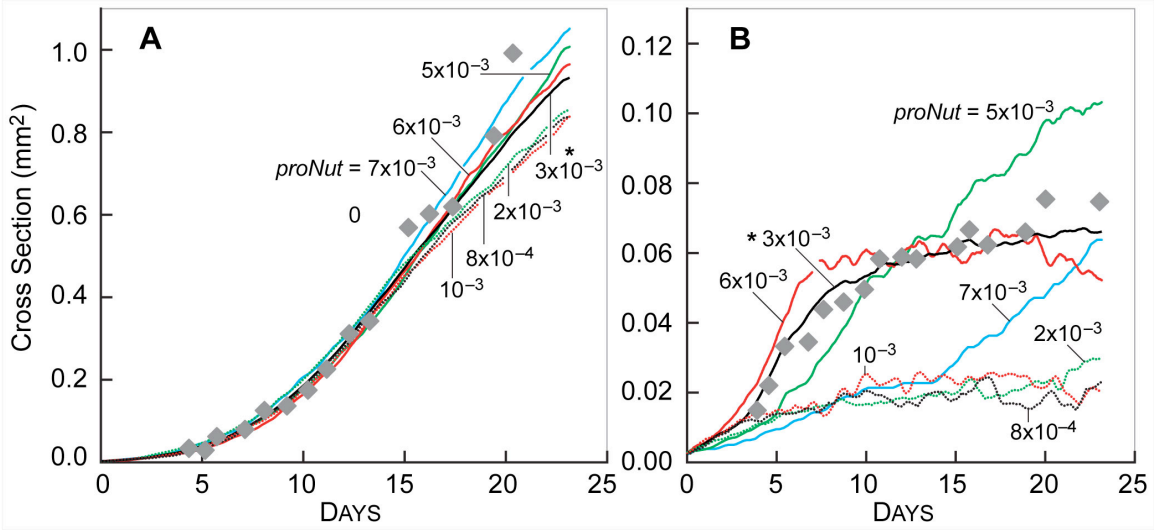


Figure 2.15. Influence of *proNut* on SMS growth.
Gray diamonds: in vitro data as in Figure 2.6. Other parameter values were those listed in Table 2.1. Colored lines are results of single experiments for the indicated values of *proNut* from 8.0 x 10^{-4} to 7.0 x 10^{-3} (same values as in Figure 2.14). (A) high NUTRIENT; (B) low NUTRIENT. *: *proNut* value in Table 2.1.

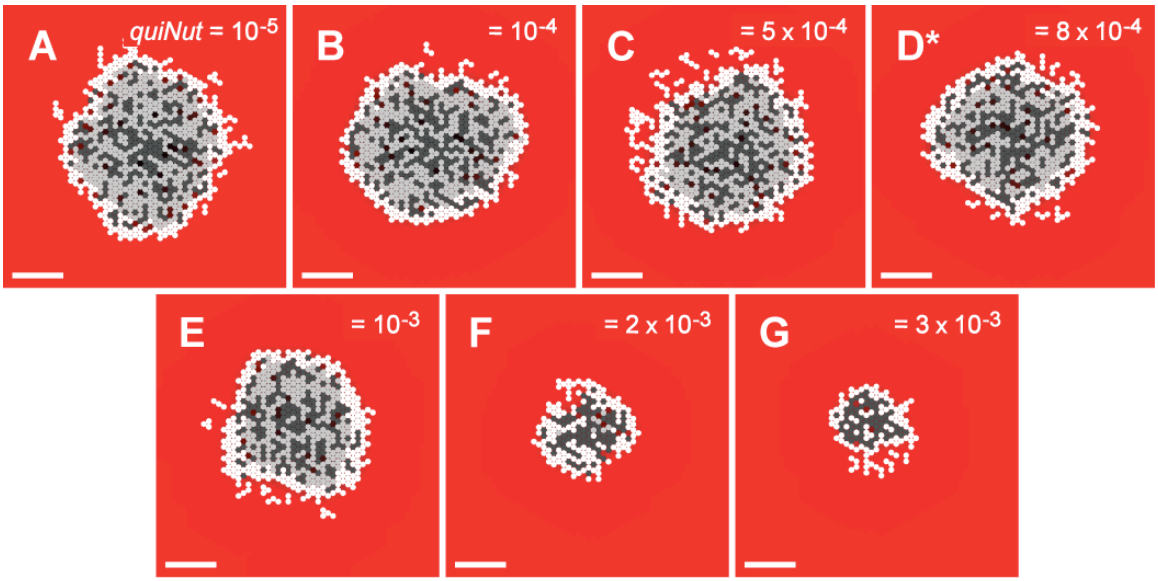


Figure 2.16. SMS cross sections at varied *quiNut* values and low NUTRIENT.
All images were recorded at 21 DAYS. Scale bar: 100 µm. Except for *quiNut*, parameter values were those listed in Table 2.1. (A)-(G): *quiNut* values are shown. Stabilization size and VIABLE rim widths decreased as *quiNut* increased, but all experiments achieved stability. *: *quiNut* value in Table 2.1.

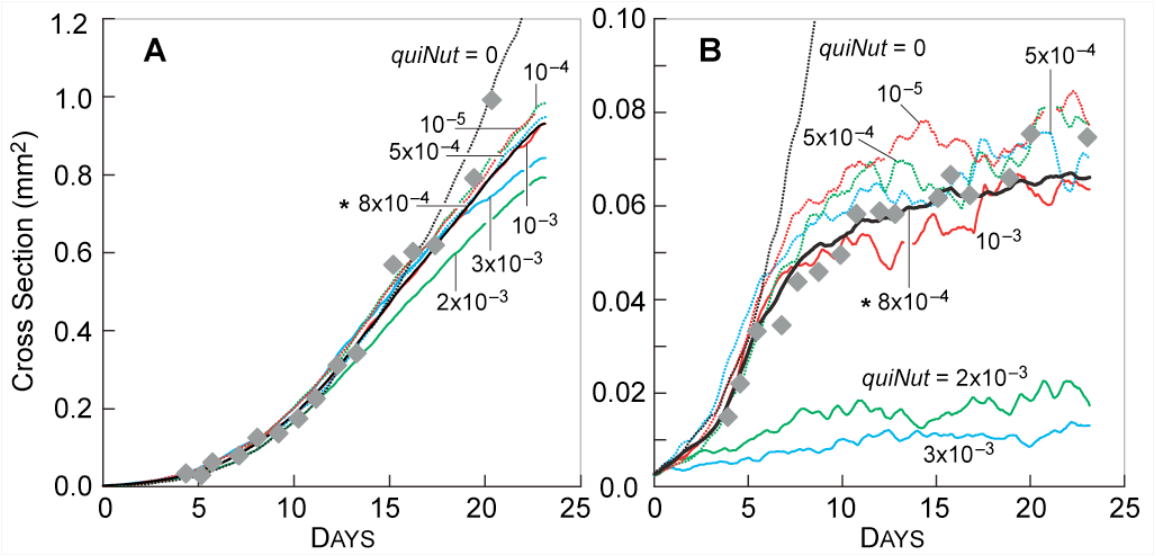


Figure 2.17. Influence of *quiNut* on SMS growth.
Gray circles: in vitro data as in Figure 2.6. Except for *quiNut*, parameter values were those listed in Table 2.1. Colored lines are results of single experiments for the indicated values of *quiNut* from 0 to 3.0 x 10^{-3} (same values as in Figure 2.17). (A) high NUTRIENT; (B) low NUTRIENT. *: *quiNut* value in Table 2.1.

The consequences of changing *proBias* at low NUTRIENT were potentially confusing, because there were two dramatic and different effects on SMS growth and morphology. The first effect, clearly observable in Figure 2.18B, was an increased growth rate when *proBias* was decreased from the default validation value of 2.25. CELLS that used smaller *proBias* values were able to create new CELLS even when their stress was higher. As a result, they created new CELLS more frequently, increasing the overall growth rates. The increase in growth rate confounded our ability to analyze the changing stability of the SMS, which was the second major effect of changing *proBias*. At very low values of *proBias,* individual CELLS proliferated quickly, and the SMS grew to the edge of the space within a few DAYS (images not shown). Consequently, it was not possible to determine whether the SMS was more or less stable at these values without adjusting other parameters, such as *prolifDelay.* At values of 1.0 and 1.25, the growth rate decreased, and the SMS was relatively unstable. As *proBias* increased further (1.5,

2.0, 2.25), the SMS became stable once again, though the growth rate decreased as the value was raised further (2.5, 3.0). We believe that if *prolifDelay* were adjusted in concert with *proBias*, maintaining the same initial doubling rate, the SMS would become unstable at low values of *proBias*. At high NUTRIENT levels, shown in Figure 2.18, the SMS growth rate decreased as *proBias* increased, but the relative SMS stability did not change.

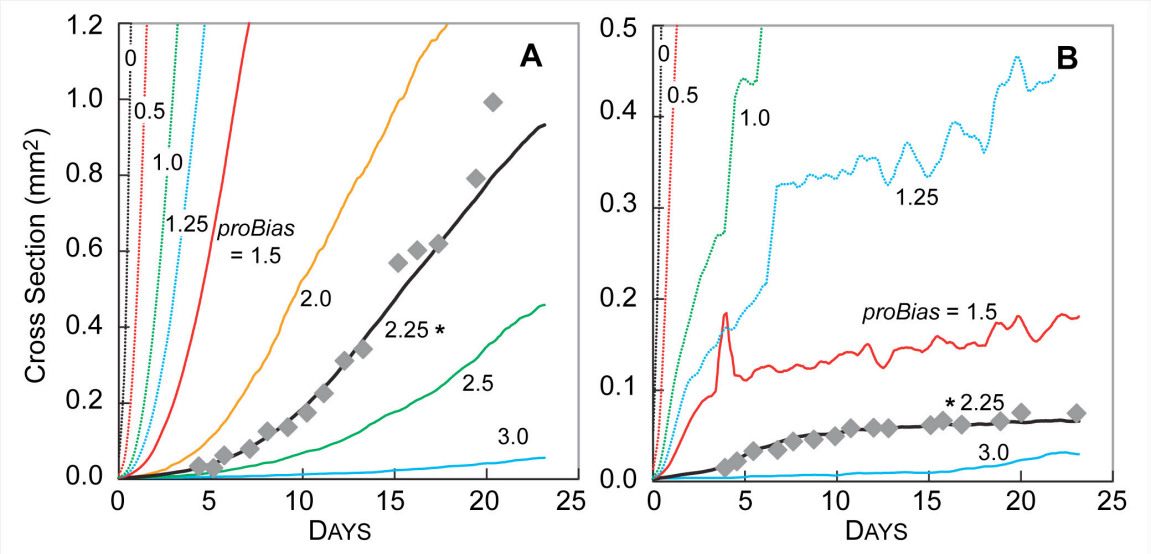


Figure 2.18. Influence of *proBias* on SMS growth.
Gray diamonds: in vitro data as in Figure 2.6. Other parameter values were those listed in Table 2.1. Colored lines are results of single experiments for the indicated values of *proBias* from 0 to 3. (A) high NUTRIENT; (B) low NUTRIENT. *: *proBias* value in Table 2.1.

2.4. Discussion

SMS events and mechanisms were not intended to be exact replicas of the actual physical or chemical events ongoing in vitro during EMT6 spheroid growth. Nor were predictions of specific events part of the intended SMS use. Rather, the intent behind our method has been, given a set of EMT6 spheroid attributes, to discover SMS computational mechanisms that might map logically and intuitively to in vitro counterparts. This has been accomplished by exploring the inverse map from phenomena to mechanism. The primary functional unit of an SMS—a CELL—does map 1:1 to an

EMT6 cell. Because an EMT6 cell is autonomous, we designed SMS CELLS to be quasi-autonomous. SMS CELLS currently have no internal components. As atomic software objects, they needed operating principles to function. Most of the principles that cause an EMT6 cell to act in a particular way when faced with specific circumstances in culture were unknown. Consequently, we needed to discover and implement operating principles that each SMS would use, evaluate those mechanisms through simulation and observation, and modify them based on the results. Following [5, 7], CELL operating principles were formulated as AXIOMS. Their specifications were tightly guided by available knowledge of EMT6 behaviors in culture [22, 34-36]. By iteratively following the diagram in Figure 1.1, we narrowed and refined early candidate AXIOMS to nine. These AXIOMS were refined further so that measures of SMS growth characteristics would match prespecified, iteratively expanded, targeted sets of EMT6 spheroid growth characteristics according to specific similarity measures. Having achieved that objective, we suggest that the resulting SMS operating principles (Table 2.2) can stand as an abstract representation of EMT6 operating principles under comparable growth conditions. We posit that the larger the targeted set of EMT6 attributes satisfactorily matched, the more realistic the mapping between SMS and EMT6 operating principles.

It is significant that within a simulation cycle one CELL can apply more than one AXIOM. This reflects the complexity inherent in even the simplest interpretation of a biological system. The amount of nutrients or growth factors in the environment, for example, can be independent of whether a cell is surrounded by other cells or isolated.

Use of AXIOM 8 in combination with the others resulted in an extreme degree of contact inhibition: CELLS that were surrounded by other CELLS did not create new CELLS.

That was a purposeful simplification. Nevertheless, the targeted attributes were achieved. The evidence indicates that some cell proliferation does occur throughout EMT6 spheroids [28], but that the frequency decreases dramatically with distance from the surface. If those observations were to be added to the list of targeted attributes, it would falsify the current SMS. Validation against that expanded attribute set would require increasing SMS complexity, possibly revising, as well as extending the list of AXIOMS. Relative to the current SMS, the fraction of CELL creation events occurring at the surface would be reduced and counterbalanced by division events occurring elsewhere.

The current set of abstract, axiomatic operating principles is believed to be the source of the discrepancy between in silico and in vitro growth at high NUTRIENT (Figure 2.6). The SMS can be parameterized so that simulated growth more closely matches the higher NUTRIENT data (not shown), but at the expense of achieving a much poorer match to the low NUTRIENT data. Note that the differences in growth properties at low and high concentrations of oxygen and glucose are more extreme for the referent data than is seen with other available sets of growth data, such as the data used by [2]. Achieving a tighter match would require adding more detail.

Whereas Freyer and Sutherland described the inhibitory actions of a tumor extract on proliferating cells [22], they did not separate the components to identify the source of inhibition. LaRue et al. [28] observed cyclin-dependent kinase inhibitors that are associated with cell-cycle arrest, but they did not demonstrate a causative role. Researchers have speculated that a factor in spheroid growth stabilization may be cell inhibition caused by some material being released from necrotic cells [22, 28]. We did not include such an attribute among those targeted, in part because it had not been

confirmed. Nevertheless, the current SMS successfully produced stable spheroids without the production and action of such a factor, effectively establishing that one is not required for growth stabilization at biologically realistic SMS parameter settings. Of course, we cannot conclude from this in silico evidence that a necrotic inhibitor is absent in vitro. It is instead evidence that EMT6 spheroid growth stabilization need not *require* the presence of such an inhibitor. It is also useful to contrast the modeling approach used here with that used by [18, 44]. Longo et al., having achieved some degree of satisfaction about the mechanisms implemented, focused on replication and prediction of referent results from particular AXIOMS in an exploration of the forward map from generator to phenomenon. Our approach focused on discovering appropriate AXIOMS, such as the need for a potential inhibitor or a particular arrangement of neighboring components, and which were necessary and/or sufficient. We relied on falsification to select from the plausible generators.

CELLS that experienced a high STRESS were likely to move to reduce STRESS, while CELLS experiencing low STRESS were likely to proliferate and create more CELLS. As shown in Figure 2.8, some large SMS destabilized during long-term growth. We determined that this behavior was caused by the probabilistic, local nature of the STRESS based movement and proliferation algorithms. At small SMS sizes, all deviations from the minimum-STRESS, convex curvatures are corrected by the movement and proliferation algorithms within a small number of simulation cycles. For much larger sizes, however, local curvature can be within the variability of the STRESS algorithm, yet the shape that emerges can be non-circular and irregular. That is because all AXIOM preconditions used only local information. When SMS are very large, the surface adjacent to every surface

CELL can be relatively flat (the CELLS are experiencing low STRESS), yet the overall SMS can be non-circular. If needed, the effect could be minimized in several ways, all of which would require increasing SMS complexity. The simplest for the current SMS design would be to enable sharing information about each CELL'S current STRESS with a larger cluster of neighbors.

Although other models have not explicitly controlled spheroid shape, they have nevertheless done so implicitly. For example, by placing an adhesion term in their models, Schaller et al. and Jiang et al. caused CELLS to cling together, thus minimizing surface irregularities [2, 6]. In fact, Schallar et al. noticed differences in overall shape when they used different values for the adhesion parameter. Anderson et al. found that changing the EXTRACELLULAR MATRIX structure in a simulated model of tumor invasion produced dramatic differences in tumor morphology [44]. Our analogue did not initially contain a mechanism to control SMS shape, but we found that the analogue could not mimic the targeted attributes without one. Although the SMS did not explicitly define and implement cellular adhesion like [2, 6], stress based movement and proliferation produced a similar effect.

3. Modeling MDCK epithelial morphogenesis

3.1. Introduction

Epithelial morphogenesis is fundamental to the development and functional specialization of tissues and organs. Tight regulation of tissue size, shape and polarization is critical for normal organ development and function. Disruption of these regulatory mechanisms leads to an array of diseases including autosomal dominant polycystic kidney disease, stenosis, and cancer. Epithelial cells, such as Madin-Darby canine kidney (MDCK) cells, cultured in a 3D matrix of natural basement membrane components, can recapitulate in vitro many of the in vivo growth characteristics of epithelial organs. They are thus valuable model systems for studying the cellular mechanisms of in vivo epithelial morphogenesis. Their phenotypic simplicity coupled with accumulated knowledge of their molecular biology provide excellent case studies for gleaning needed insight into how molecular events and environmental feedback pathways at subcellular levels lead to cell- and cyst-level phenotype. These model systems lend themselves to computational analysis and modeling as the means to gain that insight and improve our understanding of organogenesis.

To achieve that goal, we must first develop explanatory and easily challenged computational, mechanistic models. In biological research, explanatory mechanistic models generally precede predictive mechanistic models. The operating principles of explanatory mechanistic models of the type described herein are hypotheses about how we think phenomena are generated. The models are part of frameworks for generating and testing mechanistic hypotheses, as shown in [15, 45].

While many aspects of MDCK cyst formation are well understood, quantitative

data for cystogenesis has been lacking. The most recent computational models [5, 15, 45] relied on previously published quantitative data that described a few aspects of MDCK cyst growth in collagen cultures [46]. There is limited data available on the dynamics of cell number, cyst and lumen size, and mean cell size in Matrigel cultures. That caused previous models to assume that cell size remains constant. The presented data demonstrate that cell size varies during the course of cyst growth.

An objective of the project was to couple in vitro and in silico model systems to achieve a deeper understanding of cell behavior during MDCK cystogenesis within 3D Matrigel cultures. Of specific interest were the roles played by, and the timing of polarization, apoptosis, and lumen expansion. In order to improve our understanding of the link between individual cell behavior and cystogenesis, we proceeded in parallel on two fronts. We undertook new in vitro experiments designed to provide a more temporally and spatially fine-grained record of cell-level events during the first ten days of MDCK cystogenesis. These experiments and their results are described in this report. A thorough quantitative analysis of these results revealed a third stage of cyst growth after cyst initiation and lumen creation and expansion. That stage was characterized by the presence of a new cell state marked by a decrease in cell division rate and a cessation of the decrease in cell size observed in previous stages. We refer to a cell in that state as being “stabilized”.

We also developed and iteratively refined abstract, spatially fine-grained, multi-attribute, mechanistic, in silico, MDCK cell analogues (ISMAs) capable of cystogenesis. To create and validate ISMAs, we merged two modeling techniques while introducing several novel features. Following rounds of iterative mechanism refinement (including

falsification and validation), time-dependent measures of several in silico cystogenesis phenomena, including the sizes of cells, cysts, and lumens, cell number, and lumen number, became quantitatively indistinguishable from corresponding in vitro measures. The process led to two successful ISMAs that had similar operating principles but relied on different mechanistic hypotheses for how cells stabilized. In one, cells relied on information about the lumen. In the other, transition to the stabilized state was a simple timed event. Independent in vitro experiments [47], which used molecular interventions to alter the axis of cell division in two different ways, provided data that challenged ISMA mechanisms and the predictions of the cytogenic consequences of such interventions. ISMA mechanisms survived the falsification challenge: measures of cystogenesis during simulation experiments mimicking both interventions were quantitatively similar to in vitro data. This further supported our hypothesis that the cause-and-effect relationships (mechanisms) occurring within ISMAs during in silico cystogenesis (and thus their morphogenic agenda) have in vitro counterparts, both in the presence and absence of mechanistic interventions. By challenging these in silico mechanisms we better understand their in vitro cellular counterparts.

3.2. Methods

3.2.1. In vitro methods

A single cell suspension of MDCK cells was plated in duplicate on a layer of 100% Matrigel basement membrane (BD Biosciences) in the presence of 2% Matrigel in the media. Cysts were allowed to grow for the indicated duration then fixed with 4% paraformaldehyde. The cells were then stained as described in [48, 49]. Briefly, cells were stained with a monoclonal antibody against gp135/podocalyxn, and a polyclonal

antibody against b-Catenin. F-actin and nuclei were stained with Alexa-labeled phalloidin and Hoechst 33342 respectively. Each day, 20 cysts from the duplicate plates were selected at random and imaged using a Zeiss 510 laser scanning confocal microscope (Carl Zeiss Inc.). Images were acquired sequentially in four separate channels.

Cell number was determined by counting the nuclei, when visible, and actin borders when not. Cyst and lumen perimeter were traced using ImageJ and the size of the cyst and lumen within each cross section was calculated using the analyze tool. Cellular area was found by subtracting lumen area from cyst area; mean cell area was found by dividing cellular area by the number of cells; and the ratio of cellular area to cyst area was found by dividing cellular area by cyst area. Standard deviations and Similarity Measure values (defined in Results) were calculated using R. The number of lumens in each cyst was found by counting the discrete spaces within the cyst bordered by gp135/podocalyxn and actin.

The data generated by the in vitro experiments was quantitatively consistent with results from previous studies [47, 50, 51], as well as being internally consistent. The goal of conducting the in vitro experiments was to provide a particular quantitative perspective on MDCK cystogenesis. We sought an abstract mechanistic explanation of one set of cytogenic trajectories. Repeated in vitro experiments using a different batch of cells could result in distinct cytogenic trajectories, which might not be explained by the current ISMAs. Understanding and simulating such different trajectories is outside the scope of this project.

3.2.2. ISMA uses

An early task in any modeling effort is to state near- and long-term uses; one must then strive to follow a model development path intended to achieve those uses. When dealing with biology, having explanatory mechanistic models necessarily precedes having predictive mechanistic models. This project is an important, early step in developing explanatory mechanistic models of cystogenesis. A truly useful explanatory mechanistic model is one in which we can observe putative cause-effect events at several layers as they unfold. Given those considerations, we envisioned six near-term ISMA uses. 1) Instantiate and challenge hypotheses about mechanisms of cystogenesis by MDCK cells under different culture conditions. 2) Make it easy to follow mechanistic processes and trace cause-effect relationships. 3) Achieve measures of CYSTOGENESIS during ISMA executions of increasingly autonomous CELLS that are quantitatively similar to referent measures (i.e., they achieve targeted SMs). 4) Achieve increasing overlap of an MDCK cell culture's phenotype by an ISMA phenotype. 5) For validated ISMAs, explore the consequences of mechanistic interventions on measures of CYSTOGENESIS. 6) Expose possible gaps in our knowledge of MDCK cell cystogenesis. Implicit in these uses is the ability of ISMA behaviors under different conditions to stand as predictions of MDCK cell and cyst behaviors under comparable conditions.

The preceding are prerequisites for achieving six long-term ISMA uses. 1) Enable replacing ISMA operating principles with concrete mechanisms composed of interacting components. So doing is required to enable hierarchical linkage of molecular level details with specific phenotypic attributes. 2) Execute in silico experiments that test the effect on ISMA CYSTOGENESIS of simulated chemical and genetic interventions that affect CELL behaviors. 3) Enable continuous refinement of increasingly trustable, complex,

biomimetic mechanisms that stand as plausible explanations for increasingly large sets of multi-attribute, multi-source wet-lab data. 4) Represent uncertainty at multiple levels, including uncertainty in mechanistic hypotheses; provide plausible representations of sources of variability in referent data and phenomena. 5) Enable straightforward redeployment and adaptation of ISMA components to represent other cell types and their behaviors; examples include MCF-10A and primary mouse breast organoids. 6) Enable concrete translations between in vitro knowledge and epithelial diseases such as autosomal dominant polycystic kidney disease and cancer.

3.2.3. In silico methods

Components and mechanisms mapped as closely as possible to components and mechanisms in the referent system. ISMAs were composed of CELLS, LUMINAL space, and EXTRACELLULAR MATRIX. We set parameters such as the rate of CELL DIVISION and the initial size of CELLS to map to quantities within the in vitro system. Simulation began with 2-4 CELLS (to mimic the observed number of initial cells in vitro) on a 2D 100 x 100 hexagonal grid. CELLS expanded in size and divided using the CompuCell3D [52] cellular Potts model architecture and customized code. Each CELL occupied multiple locations on a hexagonal grid, thus allowing CELLS to expand, DIVIDE, change shape, and move in a realistic manner (Rejniak et al. [53] used an alternative method for enabling cell shape change). We coupled that with features of the agent-oriented modeling approach used successfully by [7, 54-56].

Each cycle, CELLS stepped through the same decision flow (Figure 3.1 and Figure 3.2); they applied the operating principles described below to change shape, DIVIDE, change state, create LUMENS, and DIE. Logic design and implementation was constrained

by the specifications in Table 3.1. Note that CELLS are atomic objects: they have no internal parts. All of their micromechanisms are in the form of axioms. Some axioms add behavior variability to ISMAs, as noted in Table 3.2.

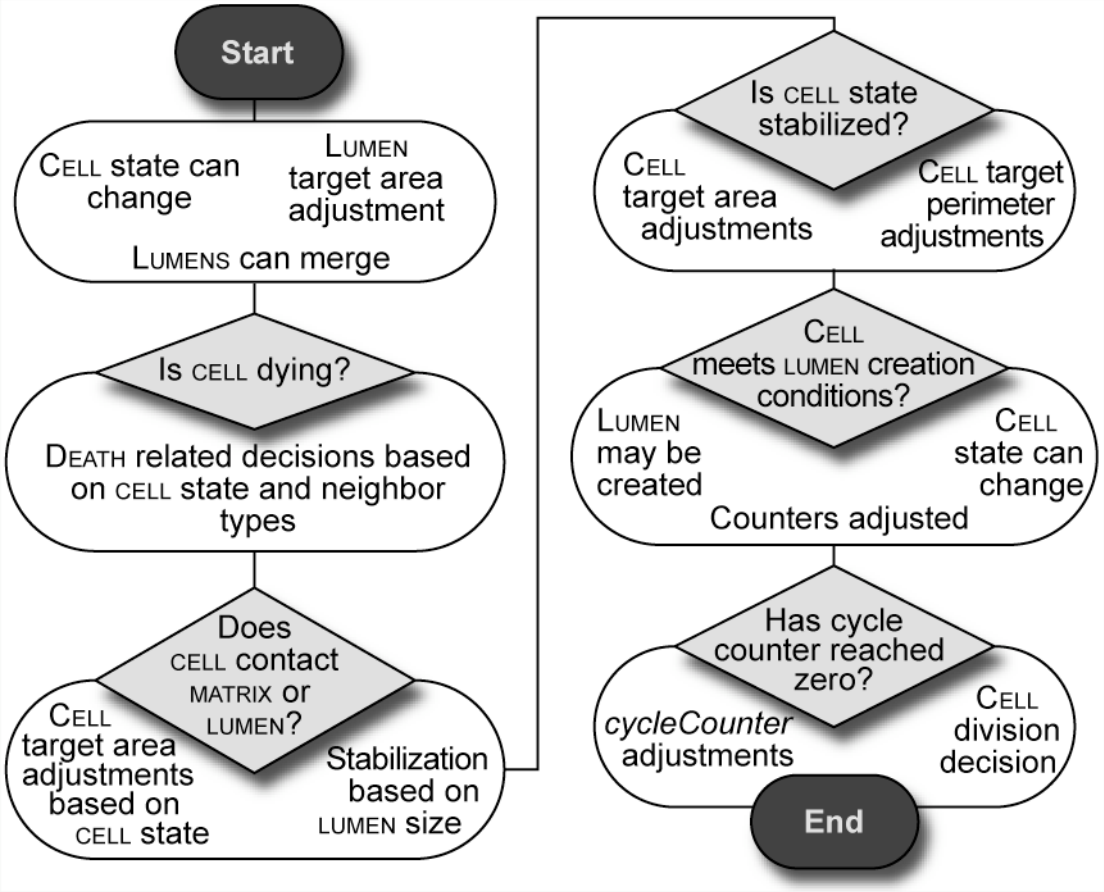


Figure 3.1. Key features of ISMA logic and decision control flow.
During a simulation cycle, each CELL steps through five logic modules sequentially to decide which actions to take based on its local environment and internal state. A LUMEN'S target area is adjusted; LUMENS can merge with each other. CELLS that are not DYING may begin to do so. CELLS adjust their area based on their state and the state of neighboring CELLS; they stabilize if the LUMEN has reached a critical size. CELLS can create new LUMENS. Under specified conditions they can divide to form new CELLS. Future versions of ISMA logic may randomize action control order to simulate the parallel nature of event occurrence both within MDCK cultures and within each cell. See Figure 3.2 for complete details of the logic within each of the five modules.

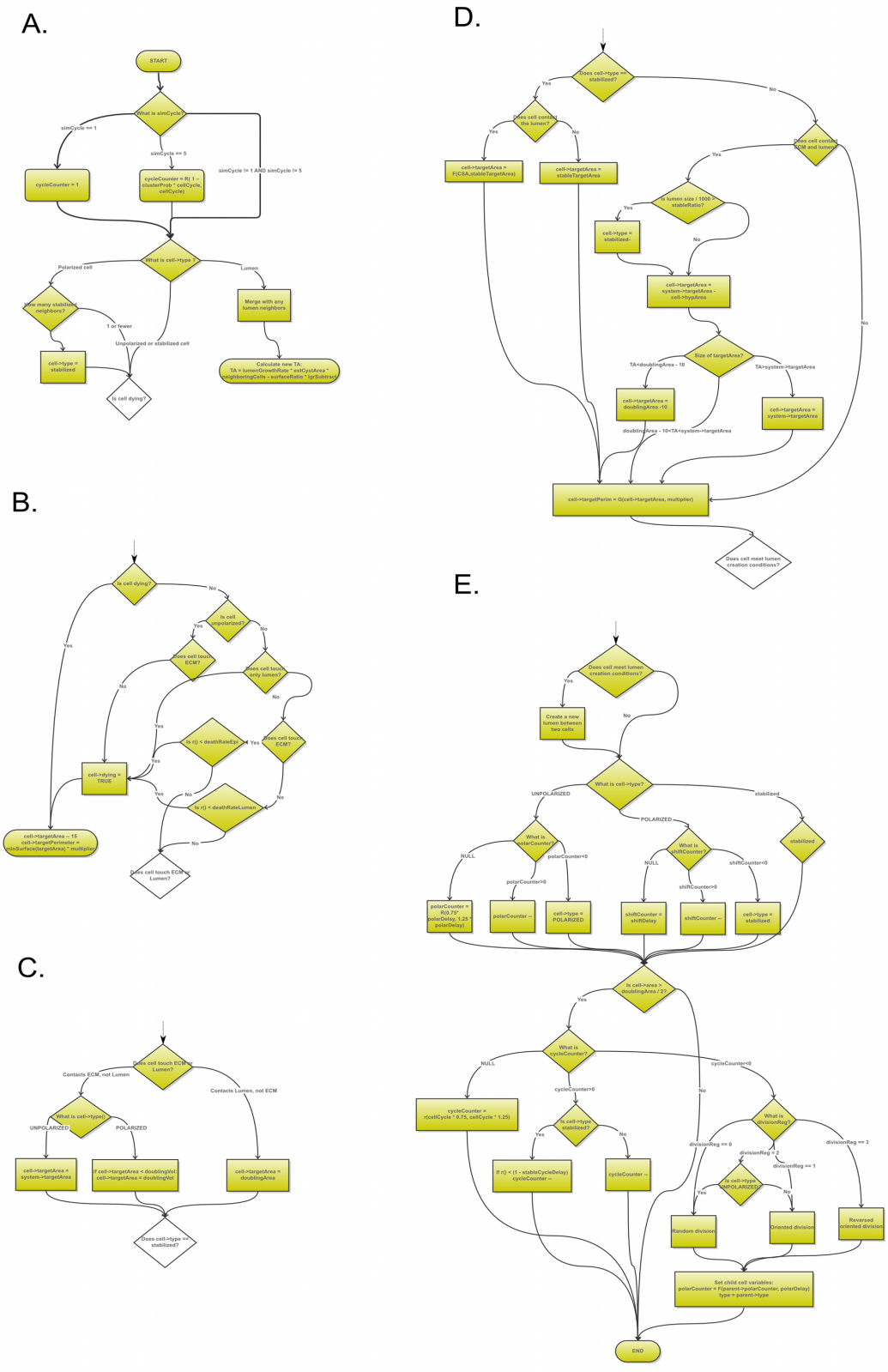


Figure 3.2. Full ISMA logic and control flow.
Shown are the details of the five components of Figure 3.1.

Except as noted, simulations ran using the parameter values in Table 3.3. A simulated DAY mapped to an in vitro day and consisted of 48 simulation cycles, equivalent to 30 minutes per cycle. Drawing on several years of prior experience experimenting on MDCK cultures, we specified that when SM1 (defined in Results) > 0.5 for nine of ten days, the results can be considered to be within the range of experimental and biological variability. Specifically, when SM1 was achieved, simulation results were taken to be experimentally indistinguishable from values obtained from an independently repeated in vitro experiment. Empirical parameter tuning was used to obtain frequencies of SLSL CYSTS comparable to that observed in vitro. When SM targets were not achieved, that specific mechanism was falsified. SMs also allowed for ISMA validation and falsification when new attributes were added to the target list (discussed below).

Table 3.1. Targeted attributes and specifications.

1.	**A:** An initial small cluster of 1-4 cells divides and increases in cell number.
	S: The ISMA begins with 2-4 CELLS, which DIVIDE after *cycleCounter* reaches 0.
2.	**A:** All cells polarize by the second day of growth.
	S: CELLS change state to POLARIZED after *polarCounter* reaches zero.
3.	**A:** One or more lumens develop by the second day of growth.
	S: CELLS within CYSTS form LUMENS after CELLS POLARIZE.
4.	**A:** A multilayer of cells separates multiple lumens.
	S: CELLS only form LUMENS when they and their neighbors do not already contact LUMEN. After a LUMEN has formed, all neighboring CELLS contact a single LUMEN.
5.	**A:** Cells can undergo apoptosis whether or not they contact the extracellular matrix.
	S: CELLS DIE with specified probability. That value is larger for CELLS not in contact with MATRIX.
6.	**A:** The increase in cell number over time is similar that shown in Figure 2, leveling off at day 6.
	S: When LUMEN size reaches a critical value, a mechanism causes CELLS to stabilize.
7.	**A:** The increase in cyst size over time is similar to that shown in Figure 2.
	S: CYST size is a function of CELL area, CELL number, and LUMEN size.
8.	**A:** The increase in lumen size over time is similar to that shown in Figure 2.
	S: LUMEN size is a function of CELL number, CYST perimeter, CELL stretch, and TIME.
9.	**A:** Mean cell area decreases over time as shown in Figure 2, and levels off at day 6.
	S: CELLS have distinct mechanisms for calculating TA before and after stabilization.
10.	**A:** The decrease in the ratio of cellular to cyst area over time is similar to that in Figure 2, decreasing faster during days 2-6.
	S: CELL area, LUMEN size, and CYST size must be measurable and if these quantities validate, then so must the ratio of CELLULAR to CYST area.
11.	**A:** The percentage of single-lumen, multiple lumen, and SLSL cysts each day is similar to that in Figure 3.
	S: When CELLS lack LUMEN contact, they can create new LUMENS. LUMEN creation occurs at the site of previous CELL DIVISION. LUMENS can expand and merge. CELLS that have stabilized cannot create a new LUMEN.
12.	**A:** The percentage of cysts with apoptotic cells each day is similar to that observed in [51].
	S: CELLS shrink after beginning to DIE. The percentage of CYSTS with DYING CELLS is calculated as in vitro.
13.	**A:** When the orientation of the cell axis of division is disrupted or reversed, the percentage of normal cysts is reduced as observed in [47].
	S: CELLS orient their axis of DIVISION toward the center of prior DIVISION or toward the center of the LUMEN. Axis orientation can be randomized and reversed.

MDCK cells and cysts are the referent. The model system is called an in silico MDCK analogue (ISMA). **A:** a targeted attribute; **S:** an ISMA specification. All listed attributes were achieved. The early version of the ISMA achieved TAs 1-4, but was falsified by the quantitative data. The refined ISMA achieved all TAs except 11, which was achieved by both the LS and the TS ISMAs.

Table 3.2. Sources of stochasticity within ISMAs.

Description	Event or variable assignment	Range / Conditional	Parameters used
CELLS set *cycleCounter* if it is null	Variable	r[0.75 x *cellCycle*, 1.25 x *cellCycle*]	*cellCycle*
At simulation cycle 5, CELLS lower *cycleCounter*	Variable	r[1 – *clusterProb* x *cellCycle*, *cellCycle*]	*cellCycle, clusterProb*
Sets initial *polarCounter*	Variable	r[0.75 x *polarDelay* 1.25 x *polarDelay*]	*polarDelay*
Child CELL sets *polarCounter*	Variable	1. pVal = *polarDelay* – parent.pCounter 2. r[0.5 x pVal, 1.5 x pVal]	*polarDelay, parent.pCounter*
CELL selects random angle of DIVISION	Variable	r(0,1) x π / 2	
Stable CELL decreases *cycleCounter*	Event	If r[0,1] < (1 – *stableCycleDelay*)	*shiftedCycleDelay*
CELL with MATRIX contact DIES	Event	If r[0,1] < *deathRateLumen*	*deathRateLumen*
Cell without MATRIX contact DIES	Event	If r[0,1] < *deathRateEpi*	*deathRateEpi*
Location changes index based on G	Event	If r[0,1] < AcceptanceProbability()	Many

The listed events or variable assignments provide behavior variability during CYSTOGENESIS. Note that index changes, which are random, contribute to many CELL-level events, including shape change and LUMEN expansion.

Table 3.3. Primary ISMA parameters.

Parameter	Description	Default value	Range used
wedgeArea	W: target area of UNPOLARIZED CELLS and ideal wedge area for POLARIZED CELLS	82 grid points	30-150
lambdaArea	Multiplier controls how quickly CELLS change size to reach their individual target areas	5 grid points	0.5-20
stableTargetArea	Target area of stabilized CELLS	48 grid points	30-150
cellCycle	Used to calculate *cycleCounter*, the number of simulation cycles before a CELL DIVIDES	70 simulation cycles	20-100
lambdaPerim	Multiplier controlling how quickly CELLS change size to reach their target perimeter	2.5	0.5-10
polarDelay	Used to calculate *polarCounter*, the number of simulation cycles elapsing before an UNPOLARIZED CELL POLARIZES	42 simulation cycles	0-400
shiftDelay	In the TS ISMA, used to calculate *shiftCounter*, the number of simulation cycles elapsing before a POLARIZED CELL stabilizes	140,000 simulation cycles	0-300
doublingArea	When divided by 2, the minimal area a CELL must have to DIVIDE	41 grid points	20-100
divisionReg	How the axis of DIVISION is calculated	1	0, 1, 2, 3
multiplier	Used to calculate target perimeter of CELLS	0.6	0-1
lumenGrowthRate	Multiplier controlling rate of LUMEN expansion	0.003	0-1
deathRateLumen	Likelihood of CELLS to DIE when not touching MATRIX	0.02	0-1
deathRateEpi	Likelihood of CELLS to DIE when touching MATRIX	0.0004	0-1
clusterProb	Probability initial two CELLS will set *cycleCounter* to zero at simulation cycle 1	0.8	0-1
lgrSubtract	Multiplied by CELL stretch to reduce LUMEN expansion	27	0-300
dyingShrinkRate	Amount subtracted from target area of DYING CELLS each simulation cycle	9 grid points	0-100
stableRatio	Critical LUMEN size (multiplied by 1000) at which CELLS will stabilize	0.5 grid points	0.1-1
stableCycleDelay	(1 – x) = probability a stabilized CELL will decrement *cycleCounter*	0.85	0-1

Parameters critical to the operation of the ISMA are listed along with descriptions, default value used for simulation, and the range of values explored. To switch between the LS ISMA and the TS ISMA the values of *shiftDelay* and *stableRatio* are changed from 140,000 and 0.5 to 200 and 1000. All units are relational (e.g., grid points instead of μM, simulation cycles instead of hours).

3.2.4. Iterative Refinement Protocol

The Iterative Refinement Protocol (IR Protocol), described in [7, 54, 57, 58], provided the foundation of our methods. Based on the results of prior experiments and literature review, we selected an initial group of qualitative attributes to target and simulate (the first few in Table 3.1). We implemented a simple ISMA that mimicked them, thus achieving an initial degree of validation. We then added new data, expanding the set of targeted attributes. So doing falsified the simple analogue. That judgment was based on observation (for qualitative attributes) and values of the prespecified SMs (for quantitative attributes). The manner in which the first analogue was falsified informed us how to develop an improved version that would survive falsification. During subsequent cycles, we added new data or features from Table 3.1 to the targeted set. So doing often resulted in falsification of the then-current ISMA. On some occasions, it was clear that an incrementally more fine-grained set of mechanisms and/or components would be needed to achieve specified SMs. On other occasions, we undertook an empirical search of parameter space to find new sets of parameter values that would reestablish validation. When that search failed, new mechanisms, sometimes more fine-grained, were developed. That iterative process ended with the attributes in Table 3.1 and the corresponding in silico specifications.

The IR Protocol has a number of benefits. Chief among them is that once an ISMA is validated against targeted data, additional data can be added and the analogue reengineered without invalidating existing mechanisms. The new data will falsify the current ISMA by design, but a successful revision will survive falsification by both new and existing data. Because in silico components and mechanisms map to their in vitro equivalents, it is often the case that only a subset of ISMA components and/or operating

principles must be modified to mimic both new and original phenomena. Examples include adding a new CELL state and replacing one axiom with two more specific axioms. Because of the networked nature of all mechanistic details, each ISMA change requires some retuning of the parameterizations of several already existing (unmodified) ISMA features.

The IR Protocol consists of the following steps: first, specify a list of targeted attributes, which forms the basis for experimental hypotheses. Devise a specification that maps in silico components and operating principles to cell culture counterparts. The operating principles are expected to enable CELLS to exhibit behavior that is closely analogous to that observed in vitro. Implement the analogue in code and execute it to deduce predictions about the in silico and in vitro system. As stated in [54], analogue execution is a form of deduction, where the behavior of the analogue follows logically from the premises embodied by its initial conditions and input data. In some cases, this deduction will yield obviously invalid results, which falsifies the current list of operating principles and prompts the modification of mechanistic hypotheses. Once the analogue cannot be falsified by data specific to the current list of targeted attributes, add one or more new, targeted attributes and repeat the IR Protocol. The process facilitates mechanism exploration, leading toward deeper insight into biological counterparts. Undertaking a series of tightly coupled in silico and in vitro experiments further increases the confidence that the results of ISMA intervention experiments can stand as useful predictions of MDCK counterparts. When there is sufficient ISMA and MDCK cystogenesis similarity, we hypothesize there is corresponding mechanistic similarity. Consequently, results of ISMA intervention experiments will stand as predictions of in

vitro phenomena following corresponding in vitro interventions. Some of those predictions will merit in vitro follow-up.

3.2.5. Agent-oriented approach

An advantage of using targeted attributes and specifications is the flexibility of their implementation. We chose to implement the ISMAs using an agent-oriented approach as explained below and described in [57], but their key aspects include object-orientation, component mapping, spatial orientation, relational grounding and striving for component autonomy. Agent-oriented models are frequently implemented using object-oriented programming techniques, which allow the designer to create individual computational objects corresponding to agents and components within the specification. Components and mechanisms are mapped to analogous components and mechanisms within the referent. So doing makes translating in vitro and in silico observations back and forth more intuitive and less complex. Individual agents can serve as analogues for in vitro components. Agents are quasi-autonomous and they possess their own internal control flow and execute actions independent of enclosing agents. Grounding is defined as the units, dimensions, and/or objects to which a variable or model constituent refers. When grounding is relational, variables, parameters, and I/O are in units defined by other model components. When grounding is absolute, variables, parameters, and I/O are in real-world units like seconds and μg/ml. One advantage of using an agent-oriented approach with relational grounding [57] is that fewer assumptions are required to create or validate the ISMA, and those that are must be clearly specified.

The ISMA contains five agents:

1. The experiment agent calls the MDCK plug-in agent and the Potts agent.

2. The MDCK plug-in agent cycles through CELL agents each simulation cycle.
3. The Potts agent executes the index change step: pseudorandom index change attempts and energy calculations.
4. The CELL agents change their state and perform other actions.
5. The screenshot agent, called in a separate thread, records a screen shot at the end of the execution of the simulation cycle.

These agents were implemented using computational objects, the most significant of which are listed below:

- *Point*: a grid location.
- CELL: contains variables like *targetArea*.
- *MCell*: contains a pointer to a CELL, a list of points that the CELL occupies, and internal variables.
- *CellInventory*: contains a list of CELLS.
- *CellField*: contains a list of all points and maps points to CELLS.
- *CellMap*: maps a CELL to each *MCell.*

3.2.6. The cellular Potts model

ISMAs were developed using the CompuCell3D (CC3D) architecture [59, 60], an implementation of the Glazier-Graner-Hogeweg [61] or cellular Potts model (CPM). A CPM "cell" is not limited to a one-to-one correspondence between objects and grid locations. The CPM extends cellular automata so that each grid location contains an index specifying which simulation object contains that location. A CPM with 100 grid locations can contain anywhere from 1 to 100 CELLS. This modification allows simulations to address CELL size, shape change, and CELL-CELL adhesion. During a simulation cycle, the Potts agent calls a pseudorandom index change algorithm that randomly selects a user-specified number of locations and evaluates whether each will remain indexed to its current CELL or change to be indexed to another CELL. If the location remains indexed to the current CELL, the grid remains unchanged. When a

location's index changes, that location and the "energy" of the system are updated.

To calculate whether a location changes index from one CELL to another, ΔG is calculated; it is the change in "energy" if that location changes its index to the new CELL. An acceptance function generates a probability *p* based on the value of ΔG, and then checks if the pseudorandom number $r[0,1] < p$. When $r < p$, the change is accepted and the location is assigned to the new CELL, and if not the change is rejected. When accepted, the energy of the system changes.

For ISMA CELLS, $\Delta G = G_{new} - G_{old}$.

It calculates the value of G_{new} and G_{old} using a Hamiltonian equation:

$G_i = EnergySurface_i + EnergyPerimeter_i + EnergyAdhesion_i + EnergyConnectivity_i$.

Each of these terms is calculated through a separate equation, detailed below.

3.2.7. Surface area and perimeter

The energy calculation for *EnergySurface* depends on *LambdaArea* (λ_A) and the difference between the target surface area (TA) and the current surface area (A):

$EnergySurface = \lambda_A \times (A - TA)^2$

The larger *LambdaArea* is the more changes in TA will affect the overall energy of the system and the faster these changes will be reconciled. *LambdaArea* for CELLS is a user-set parameter, while for LUMEN it is fixed at 20 to represent the large outward force of the expanding lumen.

The calculation of *EnergyPerimeter* is similar:

$EnergyPerimeter = \lambda_P \times (P - TP)^2$

3.2.8. Adhesion, connectivity, and TIGHT JUNCTIONS

The "energy" of adhesion depends on the CELL type and its location. For location

(i, j), the energy is the sum of values calculated between (i, j) and all neighboring points residing in separate CELLS. If, for example, two of the six neighboring points reside in another CELL, then the energy of adhesion would be $2 \cdot X_{1\text{-}2}$, where $X_{1\text{-}2}$ is a parameter controlling the adhesion energy between CELLS of type 1 and type 2. Separate adhesion energy parameters are specified for each pair of CELL types (Table 3.4).

The "energy" of connectivity is generally 0, but if changing the CELL index of a location results in a location being isolated from the rest of the CELL, an energy penalty is assessed by setting *EnergyConnectivity* to be very large. As a result, CELLS cannot split into pieces except when they undergo CELL DIVISION.

In addition to maintaining connectivity between all points in a CELL, an ISMA maintains integrity between TIGHT JUNCTIONS, preventing them from being remodeled in the index change step during a simulation cycle. If the ISMA detects that the change in a point would result in a TIGHT JUNCTION being remodeled, it assesses an energy penalty by setting $EnergyConnectivity_i$ to be very large. A detailed explanation of TIGHT JUNCTION remodeling is provided in below.

Table 3.4. Additional ISMA parameters.

System parameters	Description	Current value	Practical range
simNumber	Simulation number		0-N
folderName	Location of data stored for this experiment		0-N
dimX	Size of simulation window in X direction	100	50-300
dimY	Size of simulation window in Y direction	100	50-300
anneal	Number of post-ending simulation cycles	50	10-100
steps	Number of simulation cycles in ISMA run	481	200-1000
temperature	Likelihood an index change will be accepted	30	5-100
flip2DimRatio	Number of index changes attempted is equal to *dimx * dimy * flip2dimratio*	3	1-10
neighborOrder	Radius of neighbor grid used to calculate energy for index change attempts	1	1-3
spindleRandom	Random value added to axis of DIVISION when CELLS DIVIDE	0	0-100
Adhesion parameters			
polarcell_polarcell	Energy generated by this type of interaction. Higher energy is less favorable to contact.	5	$0\text{-}10^5$
polarcell_cell	As above	20	$0\text{-}10^5$
polarcell_lumen	As above	120	$0\text{-}10^5$
polarcell_matrix	As above	40	$0\text{-}10^5$
polarcell_stablecell	As above	5	$0\text{-}10^5$
cell_cell	As above	20	$0\text{-}10^5$
cell_lumen	As above	$5x10^5$	$0\text{-}10^5$
cell_matrix	As above	15	$0\text{-}10^5$
cell_stablecell	As above	20	$0\text{-}10^5$
lumen_lumen	As above	150	$0\text{-}10^5$
lumen_matrix	As above	$5x10^5$	$0\text{-}10^5$
lumen_stablecell	As above	120	$0\text{-}10^5$
matrix_matrix	As above	150	$0\text{-}10^5$
matrix_stablecell	As above	40	$0\text{-}10^5$
stablecell_stablecell	As above	5	$0\text{-}10^5$

Parameters used for the ISMA system and for the adhesion plug-in are listed along with descriptions, default values used for simulation, and parameter ranges that are expected to give normal results.

3.2.9. CompuCell3D and custom code

CC3D is designed from a system-based perspective. Each simulation cycle, each aspect of the system is executed, from the index change step that selects random points, to the plug-ins that update aspects of the system. CC3D was not designed from an agent-oriented perspective, so it was necessary to expand it to gain required capabilities. *MCell* objects were added to CELL objects to create a bi-directional mapping between individual

points and the CELLS that contained them. These objects and their control flow were executed in sequence by the MDCK plug-in to grant full agency to CELLS, which previously only executed after a location within the CELL boundary changed its index. Every simulation cycle all points in the grid are surveyed to assess which CELL they are indexed to and a reference is stored in an *MCell* object corresponding to that CELL, as shown in Figure 3.3. Thereafter that *MCell* can be queried to find out what points are located within its corresponding CELL object.

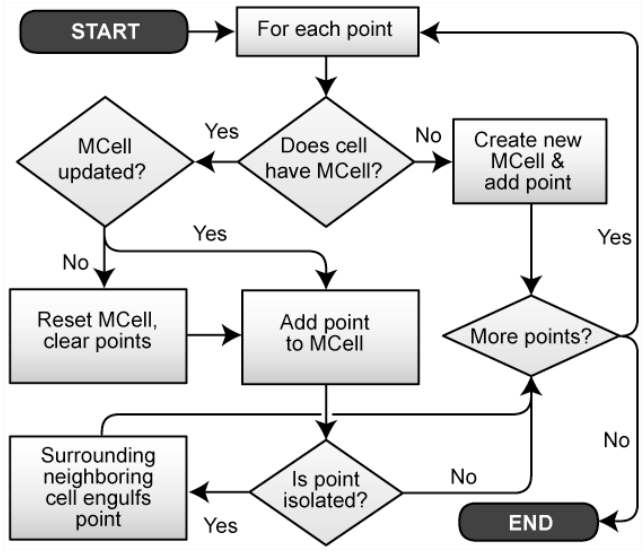


Figure 3.3. *MCell* point assignment flow chart.
An *MCell* point has no specific cystogenesis counterpart. Once per simulation cycle, each point is assigned to the *MCell* agent associated with the CELL enclosing that point. *MCell* point lists are initialized during each simulation cycle. Additionally, surrounding CELLS engulf isolated points.

The version of CompuCell3D used to develop this project has been superseded by the current available version. The capabilities provided by the current version were not judged necessary for the ISMA, especially due to the significant addition of custom code. The project was not adapted to the updated version.

3.2.10. CELLS compute their target size using a value of ideal area

As shown in Results, we observed that prior to cells stabilizing in vitro, their size correlated with the size of the cyst and its cell number. We hypothesized that operation of

yet-to-be identified micromechanisms provides each cell with a target size. We speculated that a cell might use information such as the tension between it and neighboring cells, lumen pressure, and the ratio of lumen and matrix contact area in order to update its target size. To mimic the decrease in mean cell area observed in vitro, we developed and used an algorithm that is a placeholder for yet-to-be-designed, concrete micromechanisms that can be implemented in a future ISMA. Each individual CELL adjusted its size and shape so that a target area W, the projected wedge area (a wedge that includes the portion of the perimeter in contact with MATRIX and terminates at the CYST center), would move toward or equal an ideal value. The parameter *wedgeArea* was a value based on the early (pre-stabilization) 169 μm^2 area observed in vitro. An ISMA calculated W using the following formula:

$$W = A + \frac{M^2 A}{(L+M)(L-M)}$$

A is the area of the CELL, M is one-half the number in CELL grid edges in contact with MATRIX, and L is one-half the number in cell grid edges in contact with LUMEN. This formula assumes that CYSTS are somewhat circular. Variations in actual CELL size caused by non-circular CYSTS resulted in variance in CELL area similar to that observed in vitro. The CELL subtracted W from *wedgeArea* and set its target change in area to the resulting value (with a final maximum value of *wedgeArea*). Use of this algorithm during early simulation cycles caused mean CELL area to decrease and CYST area to increase, mimicking observed in vitro data.

Once CELLS stabilized, they no longer used the above equation. Instead, CELLS strove to maintain an area that increased only slightly as contact with the LUMEN

increased. We speculated that cells within cysts in vitro must maintain a minimal cell height even as they are stretched by the expanding lumen. We specified that ISMAs use a similar guideline.

3.2.11. CELLS compute a target perimeter

From in vitro observations, it seems likely that cells have genetic and environmentally imposed targets for the areas occupied by different surfaces (cell-cell interfaces, basal, and apical). We specified that 2D CELLS have a target perimeter value (TP) that is computed using the CELL's current area. For simplicity, we specified that a CELL compute TP using the perimeter P of a circle having an area A equal to its own:

$$TP = K\sqrt{4A - 1} \times multiplier$$

K is a scaling factor and *multiplier* is user-specified. Two CELLS having identical areas will have identical TP values, so if one has a larger P the difference between P and TP will also be larger, causing that CELL to move toward circularity faster.

3.2.12. CELL POLARIZATION and stabilization

The value of *polarCounter* was set to equal a pseudorandom value $r[polarDelay \cdot 0.75, polarDelay \cdot 1.25]$ when a CELL first contacted MATRIX. Thereafter, it decreased by one each simulation cycle. Upon reaching 0, CELL state changed from UNPOLARIZED to POLARIZED. Consequently, *polarCounter* is the CELL'S counterpart to a cell, having established matrix contact, changing and moving around its components in a process that ends when tight junctions have formed and the apical surface is isolated and complete.

A correlation was observed between mean cell size and the rate of cell division in vitro, but a causal link was not apparent. Individual cells may sense the area of matrix contact in part through β1-integrin signaling [49]. They may sense the area of lateral

cell-cell contact in part using catenins and cadherins [62]. That information may influence whether a cell divides or not. As stated in Discussion, tension transduced by the subapical F-actin network could allow cells to sense the size of the lumen. Such information supported our decision to use LUMEN size as a signal for CELL stabilization. Each simulation cycle, a CELL bordering MATRIX and LUMEN queried the LUMEN for its size. When that value ÷ 1000 was greater than the parameter *stableRatio*, the CELL changed to the stabilized state.

3.2.13. Cell division

Decrementing *cycleCounter* is a CELL's counterpart to moving through the phases of the cell cycle. *CycleCounter* is a variable that is initialized based on *cellCycle* (a user-specified parameter that controls the duration of the CELL cycle) and decremented thereafter. CELLS implemented the following method of CELL DIVISION. For the first CELL and for daughter CELLS after DIVISION, the value of *cycleCounter* was set to a pseudorandom value $r[0.75 \cdot cellCycle, 1.25 \cdot cellCycle]$ and then decremented in each CELL in every simulation cycle in which the CELL had an area > doublingArea / 2. When *cycleCounter* reached zero, a CELL DIVIDED (Figure 3.4), splitting its area in half on an axis, and using the parameter *divisionReg* to determine the method of calculating the axis of DIVISION.

When *divisionReg* = 0, CELLS chose the axis of DIVISION randomly. If it was 1, CELLS used oriented DIVISION, finding their axis of DIVISION as shown in Figure 3.4. CELLS recorded the location of their MIDBODY as a point. When DIVIDING, the CELL connected the MIDBODY and the centroid with a line. The CELL assigned all points above the line to a new CELL and all points below to the old CELL. It then set the MIDBODY of

the both CELLS to the centroid of the just-divided CELL. When *divisionReg* = 2, CELLS divided randomly until they reached the POLARIZED state and then used oriented DIVISION. *DivisionReg* = 3 specified reversed DIVISION, where CELLS would find the axis of DIVISION as stated above, but then add 90 degrees, reversing DIVISION orientation.

After a CELL divided the value of *cycleCounter* for both daughter CELLS was reset to a new random value as detailed above. Its value of *polarCounter* did not change. The new CELL inherited all values from the parent CELL, except *polarCounter*, which was set to $r[0.5 \cdot polarDelay, 1.5 \cdot polarDelay] - polarDelay + polarCounter(\text{parent})$. So doing made the newly created CELL have a *polarCounter* value close to but not identical to that of the parent CELL.

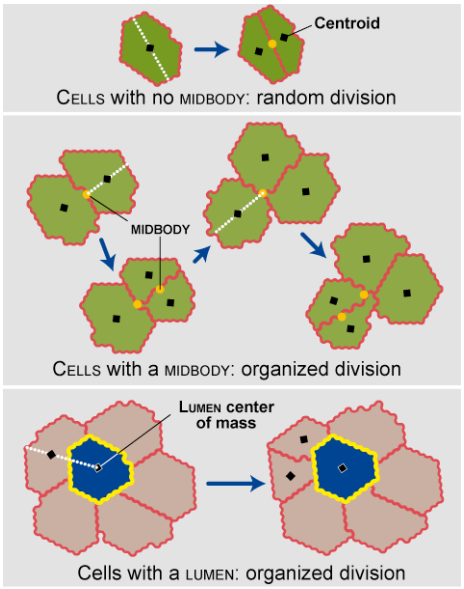


Figure 3.4. ISMA CELL DIVISION.
CELL DIVISION depends on the CELL neighborhood. Single isolated CELLS (top) that have not DIVIDED have no MIDBODY and DIVIDE with a random axis of DIVISION. When CELLS DIVIDE they find their centroid and store it as the MIDBODY of their daughter CELLS. CELLS that have previously DIVIDED and have a MIDBODY utilize it for subsequent DIVISIONS. For these organized DIVISIONS (top), the axis of DIVISION is determined by a line drawn from a CELL's current centroid to the stored MIDBODY. CELLS in contact with a LUMEN will also DIVIDE in an organized fashion (bottom), using a line between their centroid and that of the LUMEN to determine the axis of DIVISION. When the axis of DIVISION is determined, all points on one side of the line are assigned to a new CELL while all points on the other remain assigned to the original CELL.

3.2.14. CELL clustering and CELL DEATH

In vitro data analysis revealed that when the cultures began growing, the mean in vitro cell number was about 2 (Table 3.5), indicating that a small amount of clustering took place after the cells were plated. That was expected because in Matrigel culture suspended cells settle on the layer of 100% Matrigel and thus most cysts grow in the same plane. Accordingly, simulations began with a single CELL, but at simulation cycle 1, the *cycleCounter* of that CELL was reduced to 1, causing it to DIVIDE during the following simulation cycle. In addition, since in vitro cells are not always at the beginning of their cell cycle when plated, the value of *cycleCounter* for the two CELLS was changed to equal a pseudorandom number $r[(1 - clusterProb) \times cellCycle, cellCycle]$. So doing allowed the amount of clustering to be increased without changing the CELL DIVISION rate, simply by increasing *clusterProb*.

Table 3.5. Mean cell number per day for cysts grown in Matrigel.

	Projected from day 1	Projected from day 2	Projected from day 3	Projected from day 4
Day 0	2.15	2.09	2.44	2.14
Day 1	***3.35***	3.27	3.80	3.35
Day 2	5.23	***5.1***	5.93	5.22
Day 3	8.15	7.956	***9.25***	8.14
Day 4	12.72	12.41	14.43	***12.7***

Numbers in bold italic are measured, mean values and non-bold numbers are projected values. Projected values were found by multiplying or dividing the measured mean values by the scaling factor of 1.56. During the first four days of growth, the number of cells increased by a constant factor of 1.4 to 1.8 per day, with a value of 1.56 minimizing the percent error between projected and measured mean values. Using that scaling factor, the number of cells at day 0 was estimated to be 2.1, indicating that some clustering took place within the Matrigel culture. To reflect this observation, ISMAs implemented CELL clustering.

Cell death is an important factor in MDCK cystogenesis. However, it is not clear that it is required for cyst formation. In order to validate that CYSTS did not ignore or excessively rely on CELL DEATH for normal LUMEN formation, the amount of CELL DEATH observed in silico was quantified and compared to that observed in vitro. In vitro

analysis of cell death was conducted in [51]: MDCK cysts were cultured as in this report and fixed and stained with an antibody for activated caspase-3 (cleaved in apoptotic cells).

Within the ISMA, a CELL began DYING when a pseudorandom number $r[0, 1]$ was less than *deathRateEpi* if the CELL contacted MATRIX or *deathRateLumen* if it did not. Once a CELL entered the DYING state it shrank until its area reached zero. It was then removed from the simulation. Each DAY, the number of CYSTS with DYING CELLS was recorded and the percentage calculated (Chapter 3 Results). The data was separated based on whether CELLS were in contact with the MATRIX or not.

Drawing on literature evidence [63-65] and expert opinion we estimated the average time between apoptotic bodies first being visible and a dying cell breaking up into pieces to be roughly five hours. We specified that the average duration between a CELL initiating DEATH and it disappearing to be ten simulation cycles, which maps to five hours. The value of the parameter *dyingShrinkRate* specified the amount that the TA of a DYING CELL was lowered each simulation cycle (Table 3.3). Mean DYING time ranged from 6.5 to 13 simulation cycles, with an overall mean value of 9.2, which maps to 4.6 hours, when one simulation cycle is grounded to 30 minutes.

3.2.15. LUMENS and their creation

POLARIZED CELLS create a new LUMEN when two conditions are met. 1) The CELL contacts MATRIX, but is not in contact with an existing LUMEN. 2) The location chosen for LUMEN creation is adjacent to another POLARIZED CELL also not in contact with an existing LUMEN. The point chosen for LUMEN creation is the CELL'S MIDBODY (Figure 3.4), which was the centroid of the parent CELL that previously DIVIDED to create the

current CELL.

Lumen formation involves cells creating and secreting fluid. Having CELLS create and release units of LUMEN content could simulate that. One unit could correspond to a single grid space. Those units could merge with other units or with an existing LUMEN object. However, absent validation evidence for the other ISMA mechanisms, implementing such a fine-grained (somewhat complicated, multi-parameter) mechanism simply because it seems biomimetic would have been contrary to the IR Protocol's strong parsimony guideline. We took advantage of CC3D capabilities and elected to use a more abstract, simpler approach. There is no disadvantage in doing so because a strength of this class of analogues is that a simple mechanism that achieves a degree of validation can later be replaced with a more detailed and realistic counterpart. Using cross-model validation [57], this can be done without compromising other ISMA mechanisms that have also achieved degrees of validation [7].

Within ISMAs, LUMENS are a different class of "CELL" object. Their only action options are to expand and merge. After a LUMEN is created, it expands using the following axiom.

TA = *lumenGrowthRate* × *estimatedArea* × *totalNeighbors* – *lgrSubtract*

LumenGrowthRate is a user-specified parameter; *estimatedArea* is the area of CELLS in contact with the LUMEN added to the LUMEN's area; *totalNeighbors* is the number of CELLS in contact with the LUMEN; and *lgrSubtract* is a quantity based on a user-specified parameter and the degree CELLS are stretched. CELLS that are more stretched have a larger *lgrSubtract* value, reducing the rate of LUMEN expansion. A LUMEN does not have a target perimeter value—its perimeter is determined entirely by the

perimeter of the CELLS surrounding it. LUMENS can merge when their TIGHT JUNCTIONS are reorganized.

3.2.16. Tight junction maintenance

TIGHT JUNCTIONS (TJs) were implemented in order to simulate aspects of MDCK lumen expansion. TJS exist where two CELLS contact each other and a LUMEN. A TJ is two points—one in each neighboring CELL—adjacent to a point within a LUMEN (Figure 3.5). TJS control LUMEN expansion and merging, and prevent CELLS from contacting multiple LUMENS. TJs are counted around a point in two ways (Figure 3.5A). 1) If the location is within a LUMEN then the surrounding points are surveyed and any two adjacent points in different CELLS are indexed as TJs. 2) If the location is within a CELL then the surrounding points are surveyed and if any two adjacent points are in a different CELL and a LUMEN, then the point within the other CELL and the current point indexed as TJs. During index change, any index change that would change the number of TJs surrounding that location will be rejected (Figure 3.5B). Also, any index change from a CELL to a LUMEN where there are two TJs before and after the index change will be rejected. The latter axiom is included because in rare cases when a DYING CELL is shrinking it is possible for a neighboring CELL that contacts a LUMEN to come into contact with a second LUMEN without the number of TJs changing.

During the execution of individual CELL logic, LUMEN merging and expansion through TJ reorganization may occur. All indexed TJs are surveyed and the first possible merge or accepted expansion is executed. Only a single merge or expansion can be executed during a simulation cycle because after merging or expansion there are new TJs and the current TJ list becomes incorrect. The process then ends and is begun again

during the next simulation cycle.

LUMEN merging will always happen if possible. For this to occur an indexed point in a TJ must be neighboring another point within a TJ that is in contact with a different LUMEN (Figure 3.5C). If that occurs two of the points, each contacting a different LUMEN, will each change their index to that of one of the LUMENS, and then the two LUMENS, now in contact with each other, will merge to form a single LUMEN. Essentially, when four CELLS contacting two different LUMENS are in close proximity to each other then the two LUMENS will merge.

LUMEN expansion through TJ reorganization only occurs if it is energetically favorable as in the index change step detailed in the text, except that the TJ changing penalties are not assessed. For a point within a TJ to change its index into LUMEN, it first checks that no neighboring points are TJs contacting other LUMENS, that no neighboring points are contacting other LUMENS, and that no neighboring points are MATRIX or UNPOLARIZED CELLS (Figure 3.5D). Then the point will check to see if any of the neighboring points satisfy the conditions that will allow this point to change into a LUMEN. At least one neighboring point must be in a POLARIZED or stable CELL that is not already in a TJ and is not the same CELL as the current point. If this is the case the location will calculate the energy change generated by converting from CELL to LUMEN and find the resulting probability p by running the result through the acceptance function. If a pseudorandom number $r[0,1]$ is less than p the change will be accepted and the simulation updated accordingly (Figure 3.5E). This code allows LUMENS to expand as they would normally without the possibility of CELLS contacting multiple LUMENS. It also allows LUMENS separated by a single-location-wide area of CELLS to merge together

into a single LUMEN.

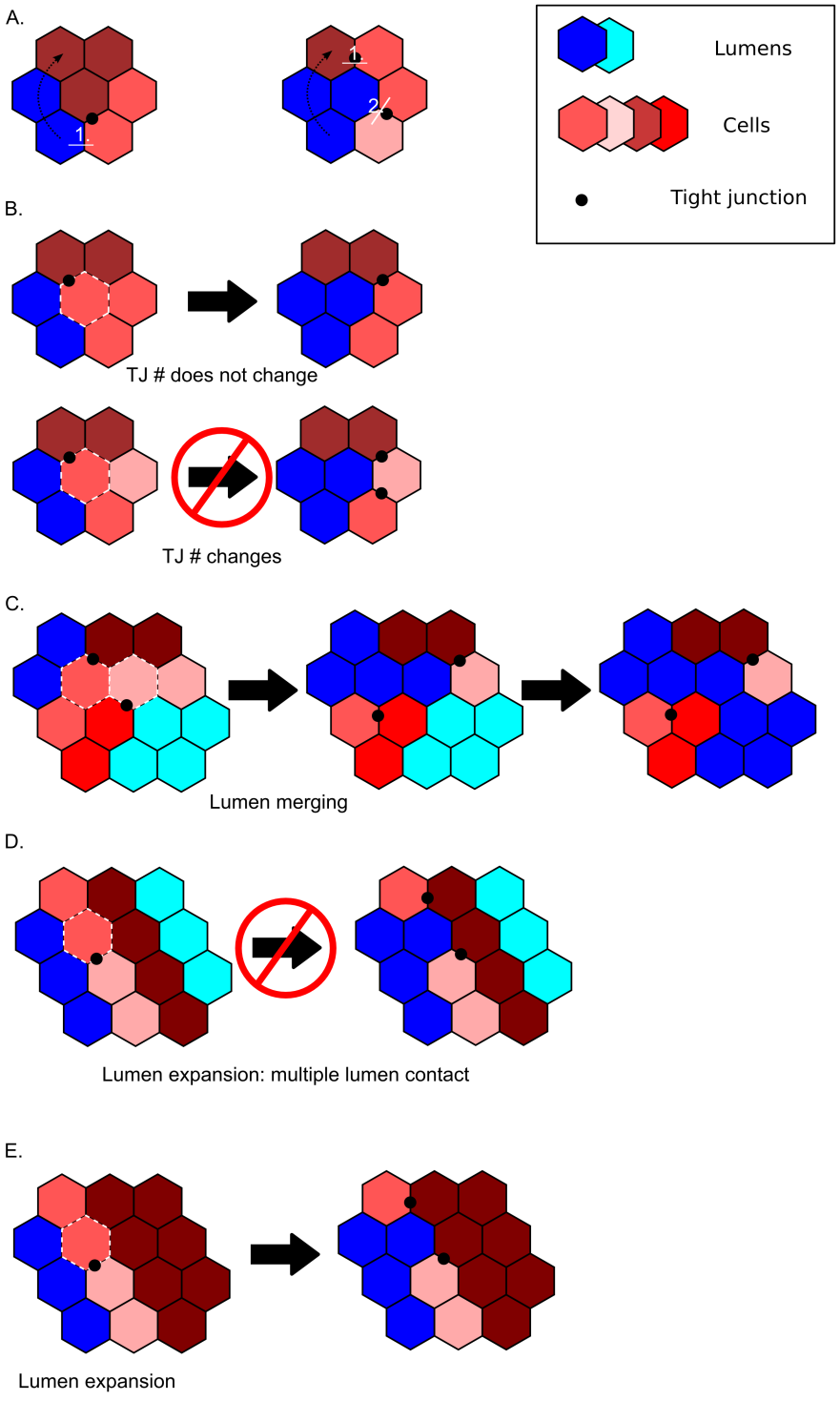


Figure 3.5. TIGHT JUNCTION reorganization.
TIGHT JUNCTIONS (TJs) prevent CELLS from contacting multiple LUMENS. A) TJ counting when a point is within a CELL (left) or LUMEN (right). B) During the index change step, index changes that result in a different number of TJs before and after the change will be rejected. C) When pairs of TJs are adjacent and meet requirements, LUMENS will merge together. D) TJ reorganization cannot occur if it will result in a CELL contacting multiple LUMENS. E) Allowed TJ reorganization results in LUMEN expansion.

3.2.17. Scaling observations from 2D to 3D

We recorded aspects of in vitro cyst growth by obtaining cross-sectional images taken through the center of cysts. These images were necessarily a 2D representation of a 3D structure. Based on the symmetry observed within the cross section, in addition to separate analysis of 3D structures, we believe that cysts were roughly symmetrical in 3D. Using this information, we extrapolated 3D values for total cell number, cyst volume, and lumen volume from the measured values of cross-sectional cell number, cyst area, and lumen area. We found that the trends observed for cell number and mean cell area held when the system was projected into 3D. If future targeted attributes required specific modeling in 3D, we could take advantage of the 3D capabilities of CompuCell3D.

3.2.18. Timed shift and geometrical mechanism ISMAs

The IR Process required the creation of intermediate ISMAs during analogue development and refinement, including a geometrical mechanism (GM) ISMA and a timed stabilization (TS) ISMA. The GM ISMA contained a number of differences in its implementation. The foremost was the method used to determine when CELLS stabilized. In the GM ISMA the following differences existed: 1) CELLS stabilized when their wedge area was more than twice their actual area instead of relying on LUMEN size. 2) CELL clustering was calculated in a different fashion, in which two random number calculations were performed instead of one. 3) LUMEN target area was calculated based on *lumenGrowthRate*, LUMEN perimeter, CELL stretch, and the number of stable CELLS

bordering the LUMEN. 4) CELLS had an increased likelihood of DYING if they did not contact the MATRIX but did contact stabilized CELLS.

The TS ISMA used an internal clock to determine when a CELL would stabilize. The internal clock was based on the variable *shiftCounter*, which was initialized to equal *shiftDelay* after a CELL POLARIZED. *ShiftCounter* was decremented at each simulation cycle, and when it reached zero the CELL would change to the stabilized state.

3.2.19. Technical specification and data storage

CC3D functions in either 2D or 3D and lets users choose between a square and hexagonal grid. It provides architecture for calculating changes in energy and accepting or rejecting changes in grid locations. The architecture is designed from a system-based perspective. Each simulation cycle, each aspect of the system is executed, from the index change step that selects random points, to the plug-ins that update aspects of the system. The CPM is straightforward and mathematically simple, making it easy to understand and its execution fast. The ISMAs were constructed using CompuCell3D 3.2.1 and custom code, as described above. Simulation code and user manual are available at <http://biosystems.ucsf.edu/research_mdckcyst.html>. All simulations were executed on a Dell Poweredge 1900 server with 2 4-core 2.33GHz 64 bit Intel Xeon processors. The system had 8 GB of RAM and a 450 GB hard drive. The system software was Ubuntu 8.04 LTS (Linux kernel 2.6). Data from simulations was captured in a MySQL 5.0 database using MySQL++ <http://tangentsoft.net/mysql++/> 3.2 to bridge between CompuCell3D <http://www.compucell3d.org/> and the database.

3.3. Results

3.3.1. Quantitative in vitro results

In order to study the process of cyst development in detail, MDCK cells were grown and observed in 3D Matrigel culture for one to ten days and analyzed quantitatively each day. As shown in Figure 3.6, cysts developed in a manner consistent with previous observations [51, 66, 67]. A suspension of mostly single MDCK cells divided to form small clusters during the first 24 hours. Most cells polarized (defined by podocalyxin localization at the nascent apical surface of the cell) during the first two days of growth and all cells polarized by day 3. Cysts developed single (11 of 20) or multiple (9 of 20) lumens by the end of day 2. Most cyst cross-sections appeared circular. The deviation from an ideal circle ranged between 2 and 5%.

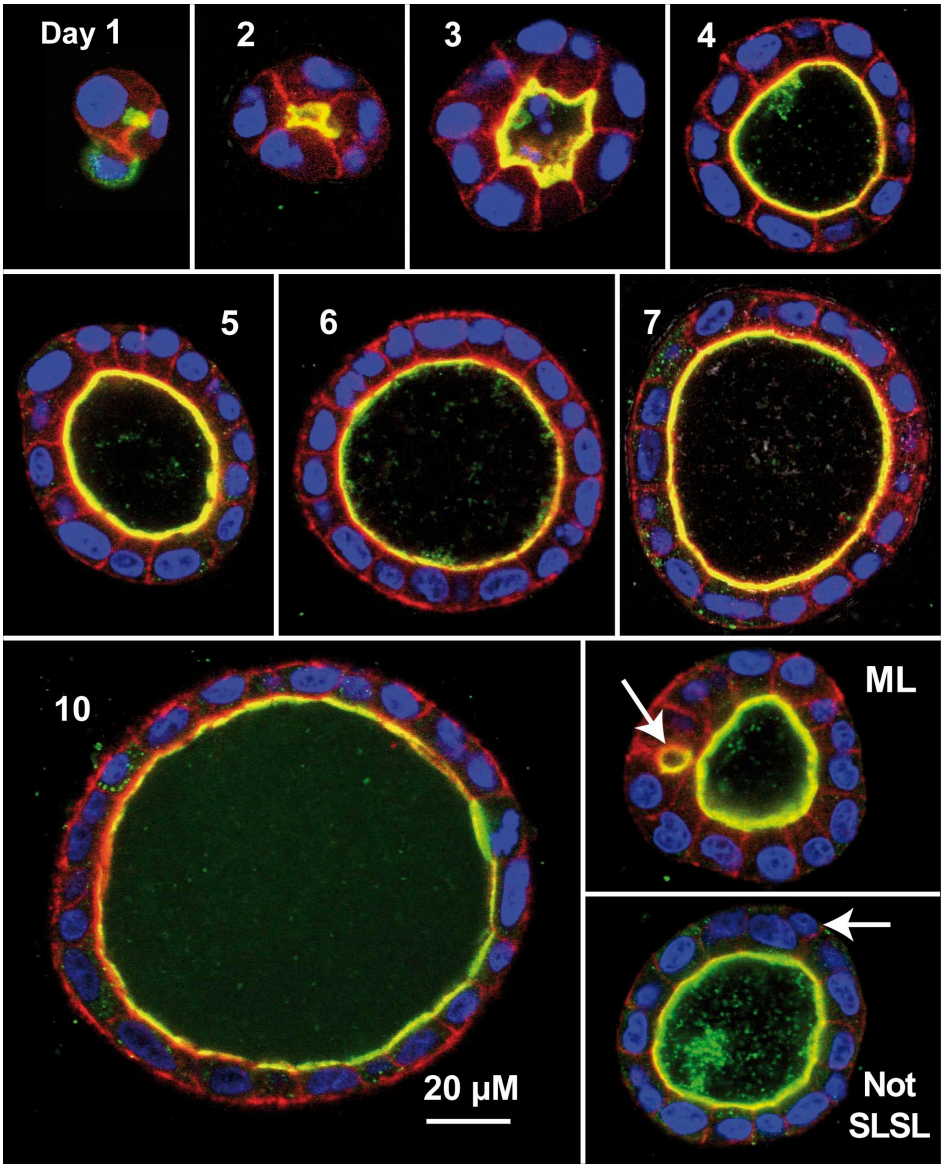


Figure 3.6. In vitro MDCK cyst cross-sections.
Culture conditions were as described in the text. Confocal images were recorded on the indicated day during cystogenesis. Colors reflect component staining as follows: red: actin; green: gp135/podocalyxn; yellow: red and green colocated; blue: nuclein; black: Matrigel.

We measured and recorded cyst and lumen area and perimeter, cell number, the number of single and multiple lumen cysts, and the number of single-lumen, single-(cell) layer (SLSL) cysts. Results are graphed in Figure 3.7 and Figure 3.8. From these results we calculated mean cell area and the ratio of total cellular area to total cyst area.

Cell number increased exponentially through day 5. It slowed and increased at a constant rate after day 6. Coincident with that shift, the variance in cell number per cyst

increased (Figure 3.7A). Cyst and lumen area increased monotonically (Figure 3.7B). Mean cell size decreased at a constant rate through day 6 (Figure 3.7C) and then leveled off at roughly the same time that cell division slowed. Mean cell size increased slightly following the shift. Cell size variance was smallest on days 5–8. We did not find a strong correlation between mean cell size and other cyst measurements, including cell number, lumen size, lumen number, or lumen perimeter/cell number. The ratio of total cellular area to cyst area (Figure 3.7D) indicated that the portion of cyst occupied by cells decreased as cysts expanded (and thus the portion occupied by lumen increased). The ratio decreased quite steeply between days 5 and 6 with very little overlap; the majority of cysts at day 5 had a ratio higher than 0.6 and the majority of cysts at day 6 had a ratio lower than 0.6. These observations taken together indicated a shift in cell behavior occurred at approximately day 5 (referred to hereafter as simply the shift). The data also supports the idea that cell compression during lumen expansion may be a factor triggering cell entry into the stabilized state.

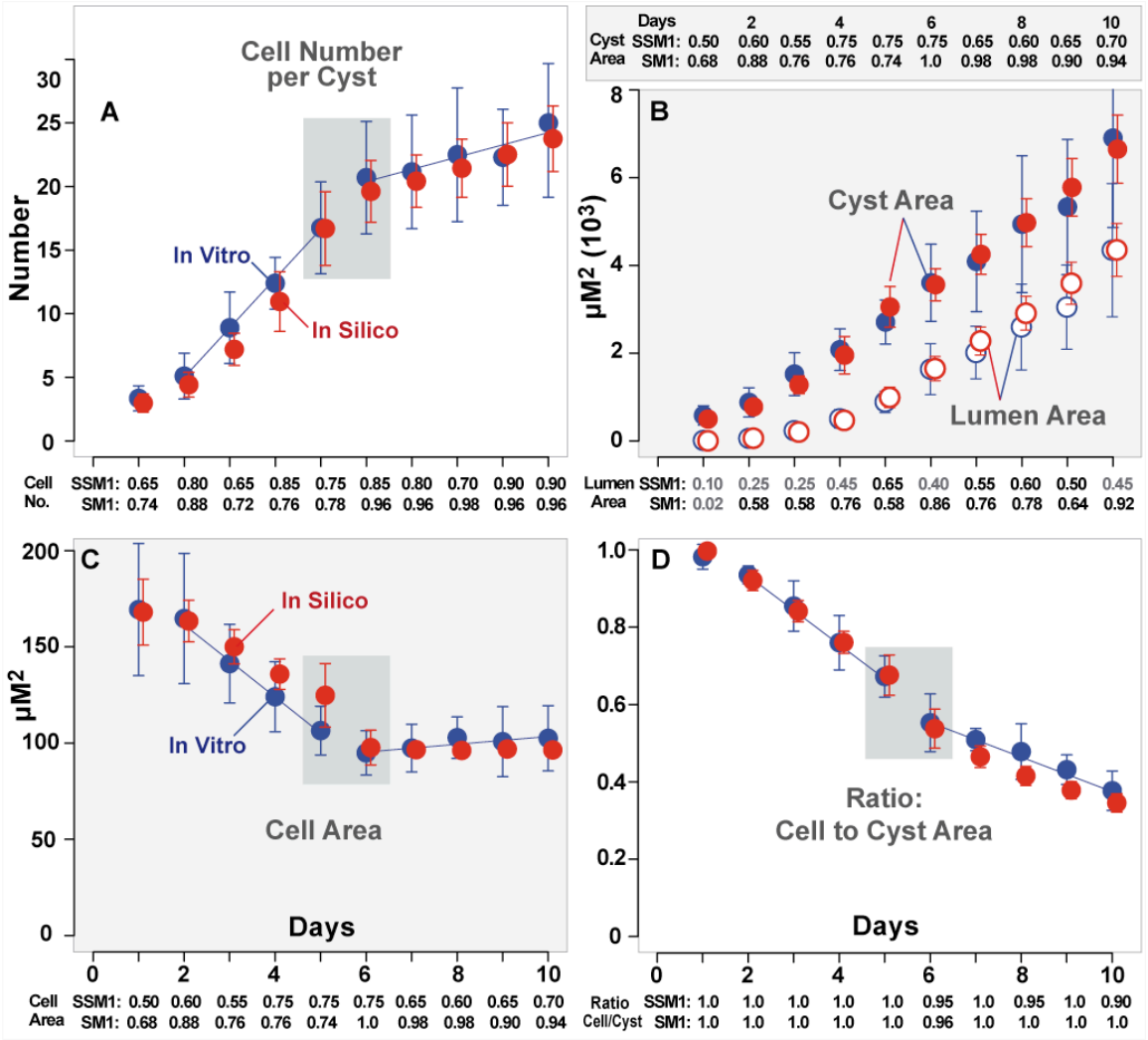


Figure 3.7. Quantitative measures of in vitro and in silico cystogenesis.
Mean values and standard deviations for (A) cell number per cyst, (B) cyst and lumen area, (C) mean individual cell area and (D) ratio: cellular to cyst area. Blue: in vitro data taken each day for ten days from 20 cysts. Red: data taken from 50 CYSTS over ten DAYS using the parameter values in Table 3.3. Gray boxes: noted changes in behavior. Blue lines: slope of in vitro growth illustrating changes in rate. SSM1: Self -Similarity Measure of in vitro growth; SSM1 indicates the percentage of in vitro values each day that fell within ± 25% of the mean in vitro value for that day. SM1: Similarity Measure for ISMA growth. SM1 indicates the percentage of ISMA values each day that fell within ± 25% of the mean in vitro value for that day. The target was that SM1 > 0.5 for nine of ten DAYS. When the target was met, we posited that ISMA measures were experimentally indistinguishable from in vitro measures. Gray SM values did not achieve targeted values.

3.3.2. Lumen percentages in vitro

During the first day of growth some cysts developed lumens, while others had no visible lumen. From days 2-10 all cysts had at least one lumen (Figure 3.8A). Multiple lumens appeared in a number of cysts, but their frequency decreased over time. Previous

studies [47] considered cysts to be “normal” if they contained a single layer of actin and apical membrane markers surrounding a single lumen. We distinguished between single-layer single-lumen (SLSL) cysts, in which all cells contact both extracellular matrix and lumen; cysts with a single lumen where some cells did not touch the extracellular matrix or the lumen; and cysts with multiple lumens (Figure 3.6). After day 2, the percentage of SLSL cysts ranged between 55% and 85% (Figure 3.8B), in rough agreement with the 80% of cysts observed by Zheng et al. to be “normal” [47]. In cases where single-lumen cysts did not have a single layer of cells, usually only one or two cells did not contact the lumen or extracellular matrix. These data indicate that the percentage of cysts with multiple lumens decreases over time, likely as smaller lumens merge together into larger. It is possible that a few cysts might increase their lumen number over time even as mean lumen number decreased, but that behavior would only be observed using time-lapse microscopy of individual cysts.

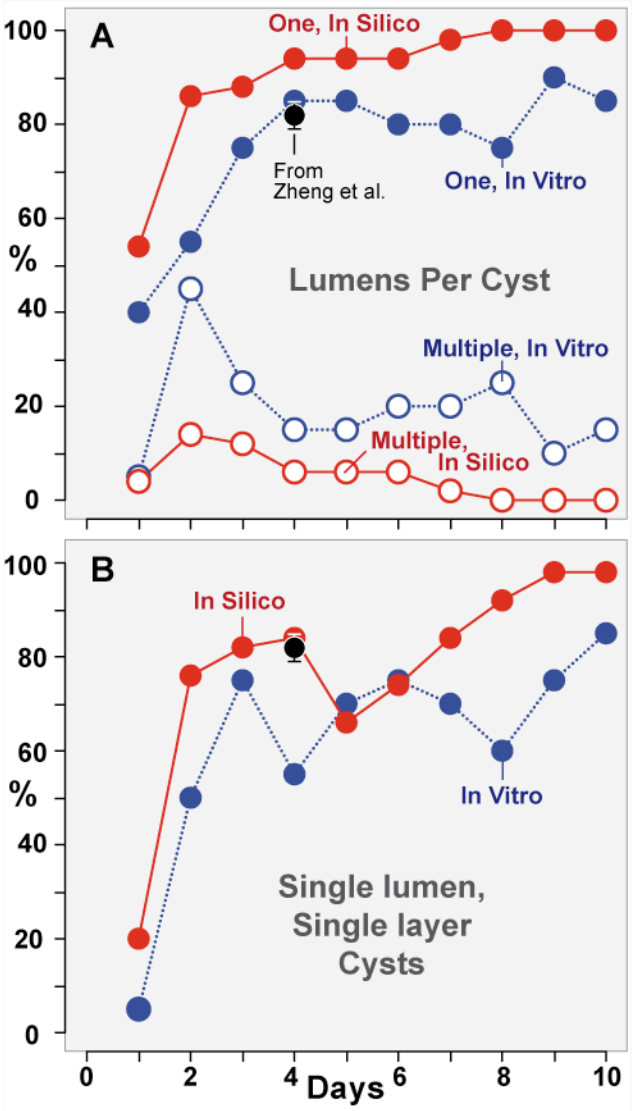


Figure 3.8. Percentage of cysts with different numbers of lumens.
(A) Percentage of cysts that have single (solid circle) or multiple (open circle) lumens. (B) Percentage of SLSL (single-layer, single-lumen) cysts. Blue: in vitro data for 20 cysts taken each day for ten days. Red: in silico data for 50 CYSTS using parameters values from Table 3.3. Black: mean and standard deviation for "normal" MDCK cysts observed by Zheng et al. [47] as described in the text. Solid lines represent continuous growth of ISMA CYSTS. Dotted lines represent discrete growth of MDCK cysts.

3.3.3. ISMAs capable of cystogenesis

In order to create and validate ISMAs, we used a number of modeling techniques and approaches, detailed in Methods. To avoid confusion between in vitro and ISMA components and mechanisms with similar names, we use SMALL CAPS when referring to the latter. Following the Iterative Refinement Protocol (IR Protocol) led to two

specifications of CELL behavior that achieved all targeted attributes in Table 3.1 and all prespecified Similarity Measures (SMs; described below). They are the lumen stabilized ISMA (LS ISMA) and the timed stabilization ISMA (TS ISMA). There are only three CELL states: UNPOLARIZED, POLARIZED, and stabilized. Both LS and TS ISMAs have a common morphogenic agenda. It is a consequence of their operating principles, which are a networked consequence of CELL state and micromechanisms. The latter are primarily axiom-dependent, and the axioms, in turn, depend on particular local and temporal conditions. The axioms are placeholders for even more fine-grained micromechanisms.

The only difference between the LS and TS ISMAs is the mechanism used by POLARIZED CELLS to shift to the stabilized state. Within the LS ISMA, POLARIZED CELLS use information about the LUMEN to decide when to stabilize. Within the TS ISMA, transition to the stabilized state is a simple timed event (each CELL used its own internal clock). We did not discover any in vitro observations that would provide a basis for selecting one micromechanism over the other.

CELL operating principles require each CELL to have knowledge of its internal state and immediate environment, including the size of the neighboring LUMEN (for the LS ISMA). CELL DIVISION is based on factors other than CELL size. Early in the process, CYST size can be independent of LUMEN size. The orientation of CELL DIVISION is extremely important in influencing the formation and number of LUMENS within a CYST.

We explored alternative mechanistic variations, but failed to find others of comparable simplicity capable of achieving all targeted attributes and prespecified SMs. For simplicity we present and discuss measures from LS ISMA simulations within the

text (Figure 3.7 and Figure 3.8) and provide the same simulation measures for TS ISMAs in Figure 3.9 and Figure 3.10. Results from earlier ISMA that were falsified because they failed to achieve one or more SMs are also presented in Figure 3.11 and Figure 3.12.

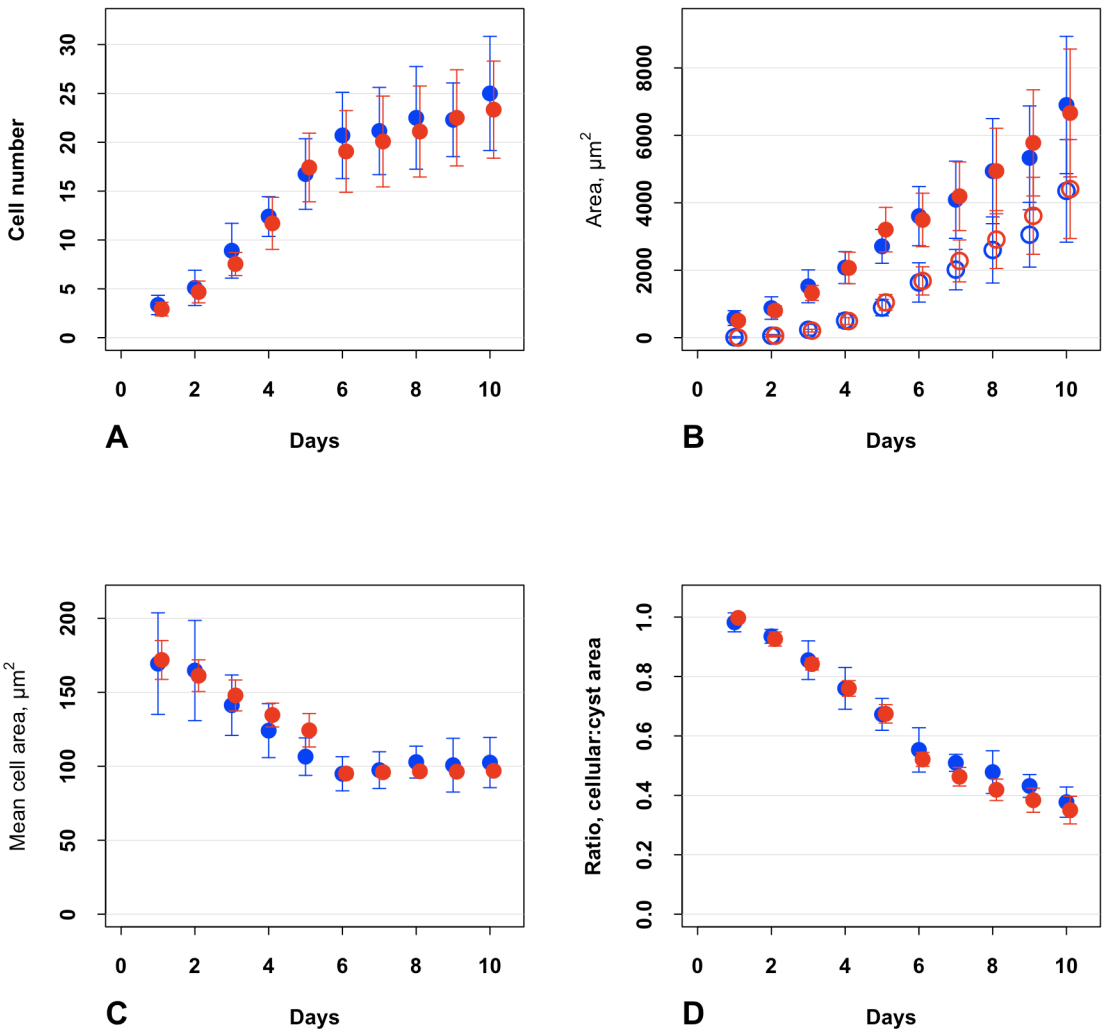


Figure 3.9. CYSTOGENESIS measures for TS ISMA.
Experiments followed the same experimental design as described in the text. Measures (red) were taken during CYSTOGENESIS. In vitro data are provided (blue) comparison**.** Designations and symbols are the same as in Figure 3.7. TS ISMA used the parameter values in Table 3.3, except for *stableRatio*, which was set to 1000 and *shiftDelay*, which was set to 200.

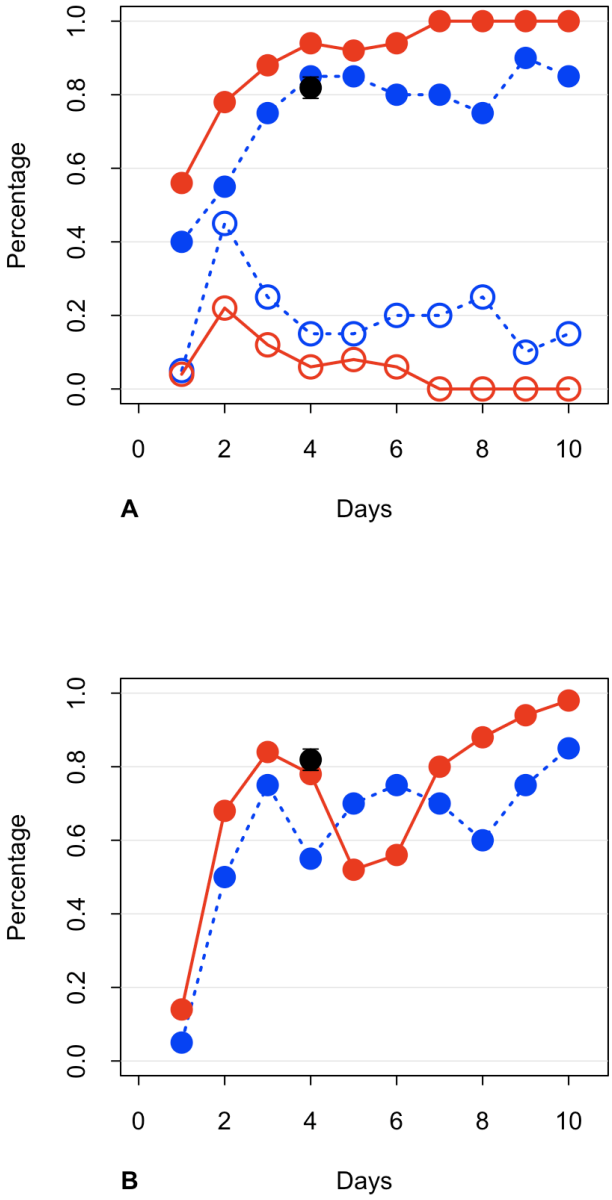


Figure 3.10. Percent of CYSTS with different numbers of lumens for TS ISMA.
The experiments are the same as in Figure 3.9. Designations and symbols are the same as in Figure 3.8.

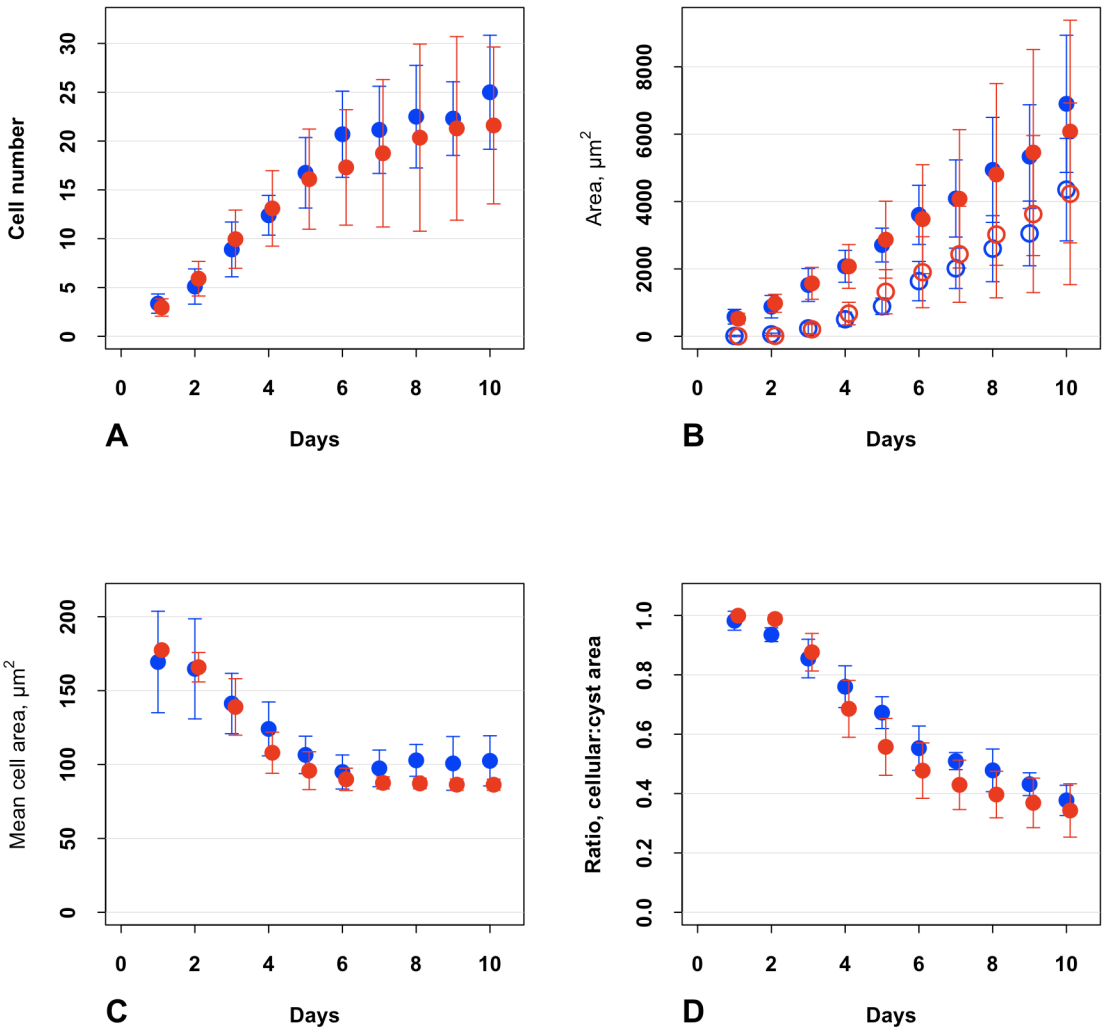


Figure 3.11. CYSTOGENESIS measures for GM ISMA.
Experiments followed the same design as in Figure 3.7 except that GM ISMAs were used. Measures, designations, and symbols are also the same as for 6. Top: note the large variances after DAY 5.

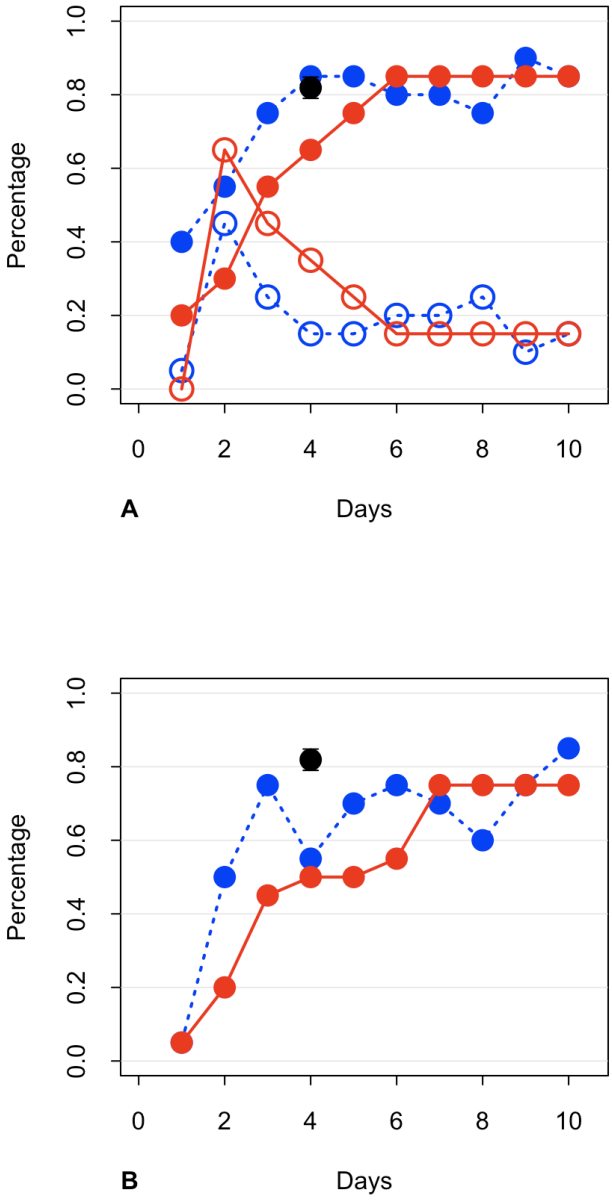


Figure 3.12. Percent of CYSTS with different numbers of LUMENS for GM ISMA.
The experiments are the same as in Figure 3.11. Designations and symbols are the same as in Figure 3.

3.3.4. Quantitative results in silico

ISMA CYSTS were similar to cysts grown within Matrigel (Figure 3.13). CYSTS began with 1-3 CELLS at DAY 0. CELLS POLARIZED and formed LUMENS within the first two DAYS (Figure 3.7 and Video S1). LUMENS and CYSTS expanded at a rate

indistinguishable from that observed in vitro. In general, a cyst formed with a single LUMEN surrounded by a single layer of POLARIZED CELLS (Figure 3.8 and Video S1). Occasionally multiple LUMENS formed, each separated by an independent layer of CELLS, such that no CELL contacted more than one LUMEN (Figure 3.13 and Video S2). The ISMA successfully achieved all qualitative and quantitative targeted attributes listed in Table 3.1.

ISMA CELL number also exhibited two growth phases, with the rate of CELL DIVISION decreasing at day 6 (Figure 3.7). LUMEN and CYST size increased at rates similar to those observed in vitro, but standard deviations were smaller. CELL size also decreased at a rate comparable to in vitro, and its standard deviations were also smaller. As indicated by the values of Similarity Measure 1 (discussed below) in Figure 3.7 and Figure 3.8, ISMAs produced quantitative results similar to in vitro values. ISMAs were executed using the parameter settings in Table 3.3, and CYST and LUMEN area were scaled by 2.25 μm^2 and perimeter by 0.75 μm.

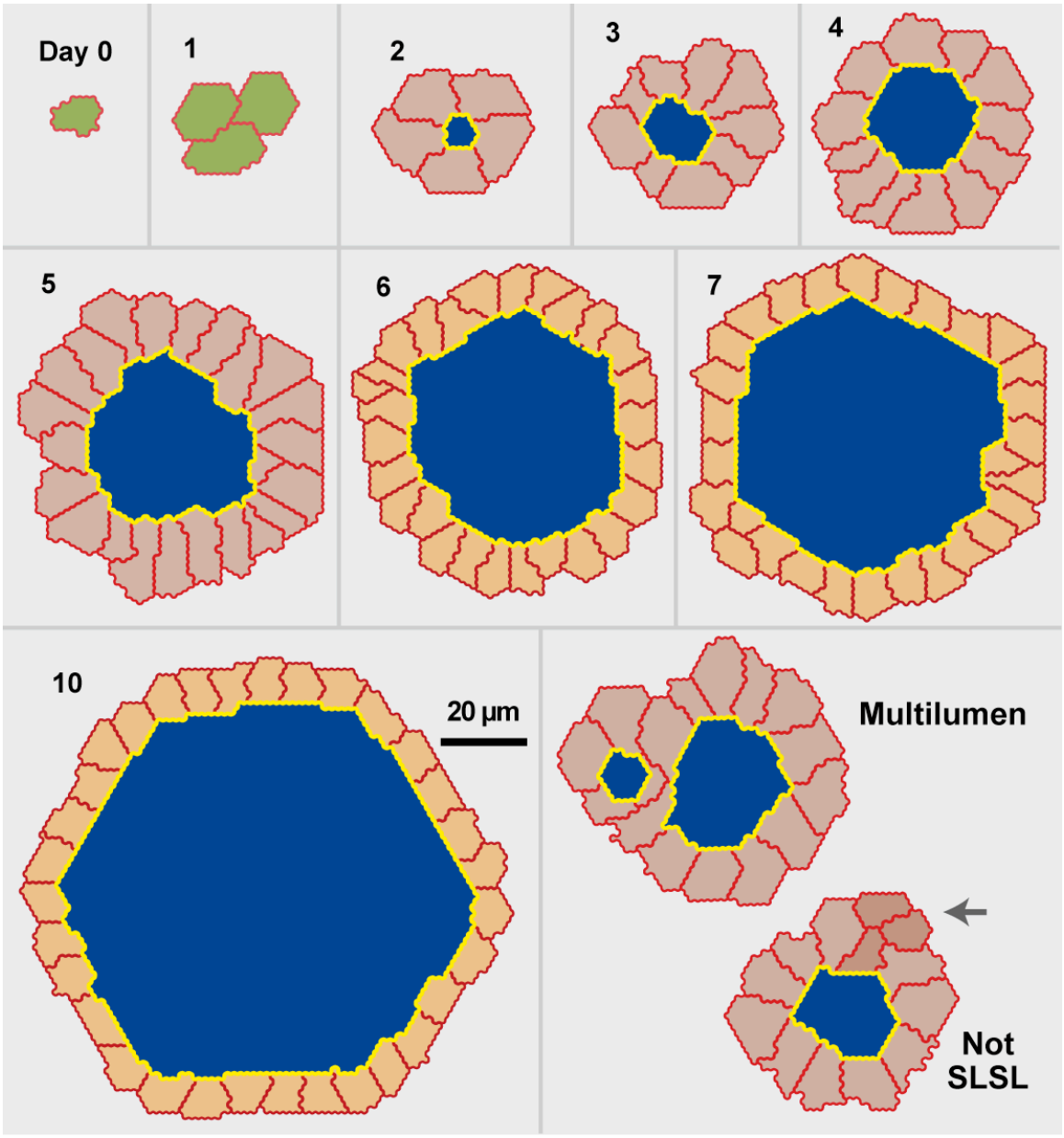


Figure 3.13. In silico MDCK analogue CYST cross sections.
Note that a regular hexagon in hexagonal space maps to a circle in continuous space. Images are from a single simulation run using parameter settings from Table 3.3. CELLS are UNPOLARIZED (green), POLARIZED (gray) or stabilized (orange). CELL-CELL and CELL-MATRIX borders are red; CELL-LUMEN borders are yellow; LUMENS are blue.

3.3.5. Lumen percentages in silico

Simulations produced single and multiple LUMEN CYSTS at frequencies comparable to those observed in vitro (Figure 3.8A), though the percentage of CYSTS with single LUMENS was slightly higher than observed in vitro. The percentage of SLSL CYSTS

(Figure 3.8B) leveled off between days 2 and 6 and then increased steadily to day 10 as LUMENS merged. CELLS that stabilized were not allowed to create new LUMENS, but could contribute to LUMEN expansion. If this restriction were to be removed and CELLS were allowed to create new LUMENS after they stabilized, the percentage of SLSL CYSTS might remain steady or decrease.

3.3.6. Similarity Measures

To provide a validation target for ISMA CYSTOGENESIS and to compare ISMA and in vitro results, we developed SMs [68], which quantified the similarity within and between the in silico and in vitro data. We posit that if in silico data satisfied the SMs, then they would be indistinguishable from data produced by a repeated in vitro experiment.

SM1 compared results from individual simulations to in vitro mean values, indicating the similarity of in silico and in vitro results. SM1 is the percentage of in silico observations that fell within ± 25% of the mean in vitro value for a given measure. SM1 values are listed in Figure 3.7. To survive falsification, > 50% of simulations must achieve the SM1 target for nine of ten DAYS, as detailed in Methods. For example, the ± 25% range for in vitro cell number at day 3 was 6.7 to 11.1 with a mean of 8.9. Seventy-two percent of simulations had CELL numbers within that range at DAY 3. SM1 values for CELL number, CYST size, mean CELL area, and the ratio of CELLULAR to CYST area exceeded 50% at all DAYS, so a degree of validation was achieved. The SM1 value for LUMEN size exceeded the 50% cutoff for nine of ten DAYS, although the values were lower.

To facilitate assessing SM1 values and comparing in vitro and in silico data, we

specified and used Self-Similarity Measure 1 (SSM1). It measured the similarity between the in vitro mean value and individual in vitro values and thus how closely grouped around the mean the individual in vitro values were. Similar to SM1, SSM1 is the percentage of individual in vitro cyst measures each day that fall within a specified range. SSM1 can be used to evaluate corresponding SM1 values. Large SSM1 values are a characteristic of measures having a small variance. Values of SSM1 were larger than the target for all measures except lumen size, indicating that lumen size in vitro varied more extensively about the mean than other quantities.

SM1 did not consider the variance of the data. To address variance, we specified SM2. It compared the coefficient of variance of in silico and in vitro experiments. SM2 measured the absolute value of the difference between the in vitro and in silico each DAY. ISMAs survived falsification if SM2 < 0.15 for nine of ten DAYS (strong validation) or < 0.25 for eight of ten DAYS (medium validation). The current ISMA achieved strong validation for CELL number, mean CELL area and the ratio of cellular to cyst area (Table 3.6). It achieved medium validation for CYST size and LUMEN size, comparable to SSM1 values.

Table 3.6. SM2 values for the ISMA.

Day	Cell number	Cyst area	Lumen area	Mean cell area	Ratio: cellular to cyst area
1	0.06	0.15	0.77	0.10	0.03
2	0.13	0.20	0.25	0.14	0.00
3	0.14	0.16	0.44	0.09	0.04
4	0.05	0.01	0.16	0.09	0.05
5	0.04	0.03	0.04	0.01	0.00
6	0.09	0.14	0.19	0.03	0.04
7	0.11	0.17	0.16	0.11	0.00
8	0.13	0.21	0.24	0.08	0.09
9	0.06	0.18	0.18	0.16	0.03
10	0.13	0.18	0.21	0.14	0.07

An SM2 value is the absolute value of the coefficient of variance (for a specific measure) subtracted from the in vitro coefficient of variance. Values over 0.25 (black) did not achieve the validation target described in the text. Values between 0.15 and 0.25 (gray) achieved the moderate validation target. Values less than 0.15 (white) achieved the strong validation target.

3.3.7. Cell death

Cell death is relatively uncommon during in vitro MDCK cyst development [51]: on a given day, no more than 15% of cysts had one or more apoptotic cells within the lumen and no more than 10% of cysts had one or more apoptotic cells with matrix contact. Cell death did occur during ISMA executions, but at slightly lower frequencies than observed in vitro (Figure 3.14). In Methods, we specified that the average duration between a cell initiating death and being removed from the simulation to be ten simulation cycles, which maps to five hours. The actual in vitro duration will affect the number of visible apoptotic cells observed each day. When we caused cells to shrink somewhat slower, the cell death values in Figure 3.14B increased. The results of such an experiment, provided in Figure 3.15, demonstrate that decreasing the value of dyingShrinkRate from 9 to 4.5 increased the mean duration of cell death (from 4.6 to 7.4 hours) and increased the percentage of dying cells. It is noteworthy that all validation targets were achieved without requiring stabilized cells to die more frequently than polarized cells. Based on current knowledge, the ISMA accurately mimics in vitro

quantitative data, but the duration of apoptosis within MDCK cells in vitro has not been quantitatively established. In order to be certain about the role played by cell death, time-lapse movies using a caspase-3-GFP will be required.

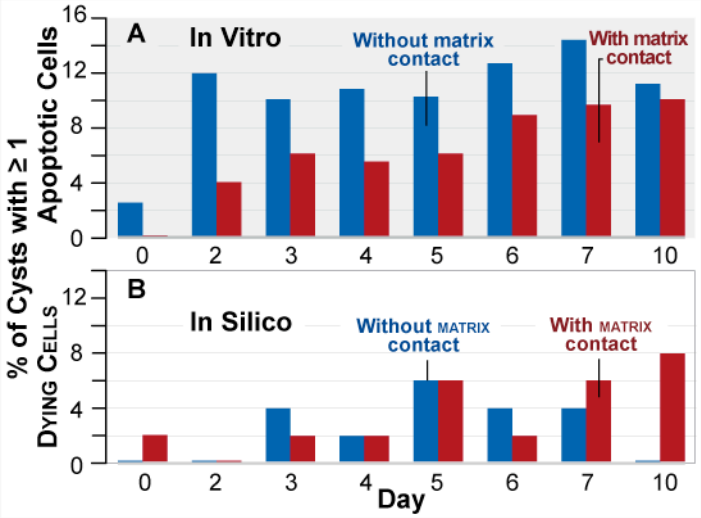


Figure 3.14. Percentage of cysts with dying cells.
(A) In vitro data reproduced from [51]. (B) ISMA data from 50 CYSTS over ten DAYS. Blue bars: percentage of cysts observed to have apoptotic cells without matrix contact. Red bars: percentage of cysts observed to have apoptotic cells with matrix contact.

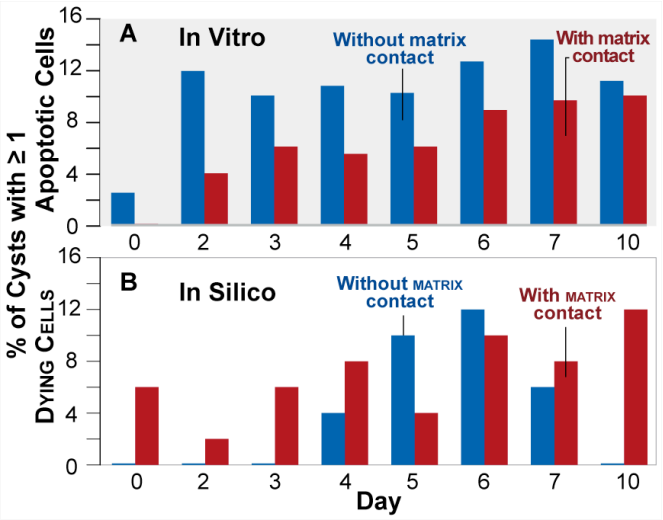


Figure 3.15. Percentage of CYSTS with DYING CELLS when *dyingShrinkRate* was reduced.
(A) In vitro data reproduced from [51]. (B) ISMA data from 50 CYSTS over ten DAYS using parameter settings from Table 3.3, except for *dyingShrinkRate*, which was changed from 9 to 4.5. Blue bars: percentage of cysts observed to have apoptotic cells without matrix contact. Red bars: percentage of cysts observed to have apoptotic cells with matrix contact.

3.3.8. Altering CELL DIVISION orientation in silico dramatically alters CYST morphology

After the ISMAs achieved the above, targeted attributes, Zheng et al. [47] reported measuring the consequences of disrupting cell division orientation on MDCK cyst morphology. Knocking down LGN, which plays a role in spindle orientation during cell division, caused cell division orientation to become random instead of aligning with the axis perpendicular to the cellular plane. The frequency of "normal" cysts decreased from roughly 80% to 20-30%. We added those observations to our targeted attributes list and then explored the degree to which CYST morphology following a comparable ISMA intervention would mimic the in vitro results, thus surviving the challenge. We altered CELL DIVISION so that all CELLS divided with a random orientation. The results (Figure 3.16A) were similar to those of Zheng et al. The altered ISMA produced less than 20% SLSL CYSTS and more than 30% multi-LUMEN CYSTS at DAYS 2 through 9. Additional details are available in Figure 3.17.

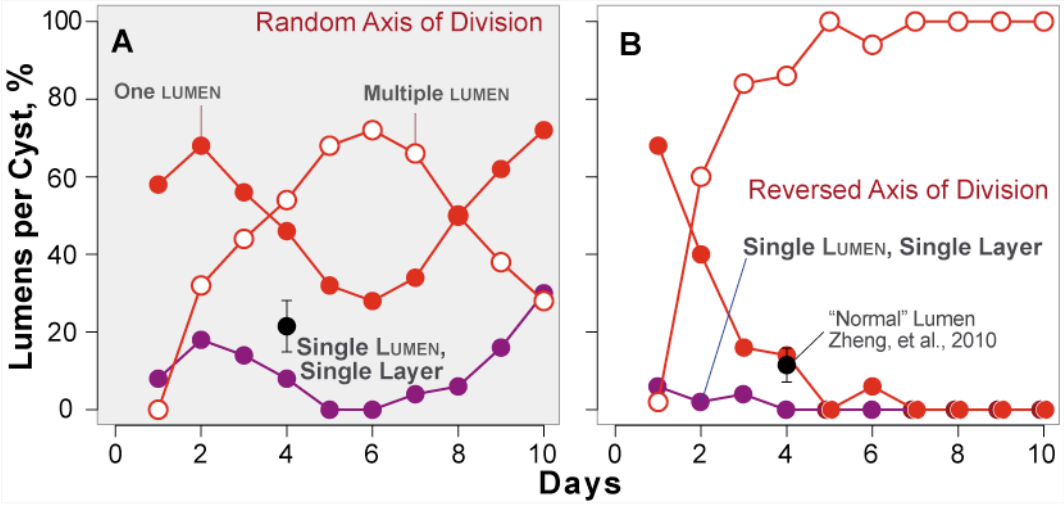


Figure 3.16. Percentage of ISMA CYSTS with varied LUMEN number when the axis of CELL DIVISION is abnormal.
Shown are the percentages of CYSTS that have single (solid red circles) or multiple (open red circles) LUMENS when the axis DIVISION is (A) random or (B) reversed (rotated 90°) along with the percentage of CYSTS that are SLSL (purple circles) when the axis of CELL DIVISION is (A) random or (B) reversed. Black (A and B): mean and standard deviation for "normal" MDCK cysts observed by Zheng et al. [47]. The in vitro control data are shown in Figure 3.8.

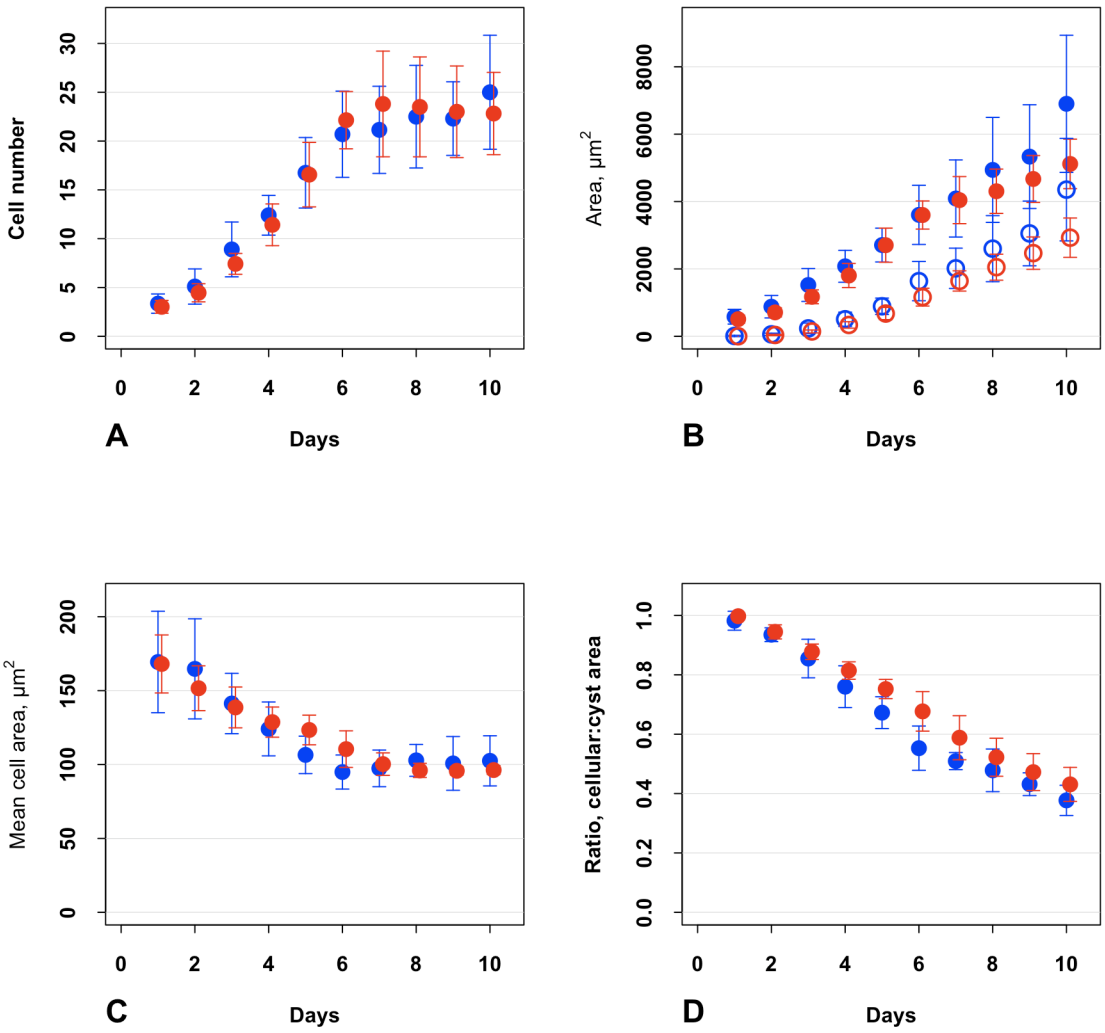


Figure 3.17. CYSTOGENESIS measures when the axis of CELL division is random.
Experiments followed the same design as in Figure 3.7. Measures, designations, and symbols are also the same. ISMAs used the parameter values in Table 3.3, except for *divisionReg*, which was set to 0.

In a second experiment, Zheng et al. targeted LGN to the apical membrane. So doing rotated the axis of division by 90°, thus reversing cell division orientation. The procedure reduced the frequency of normal cysts to roughly 10%. We conducted a similar experiment by modifying ISMAs so that the axis of DIVISION was parallel, rather than perpendicular to the LUMEN edge. That intervention produced SLSL CYSTS less than 10% of the time (Figure 3.16B and Figure 3.18). ISMAs survived both challenges; in

both cases, altering the orientation of CELL DIVISION decreased the percentage of single LUMEN and SLSL CYSTS to a degree similar to that observed within in vitro experiments.

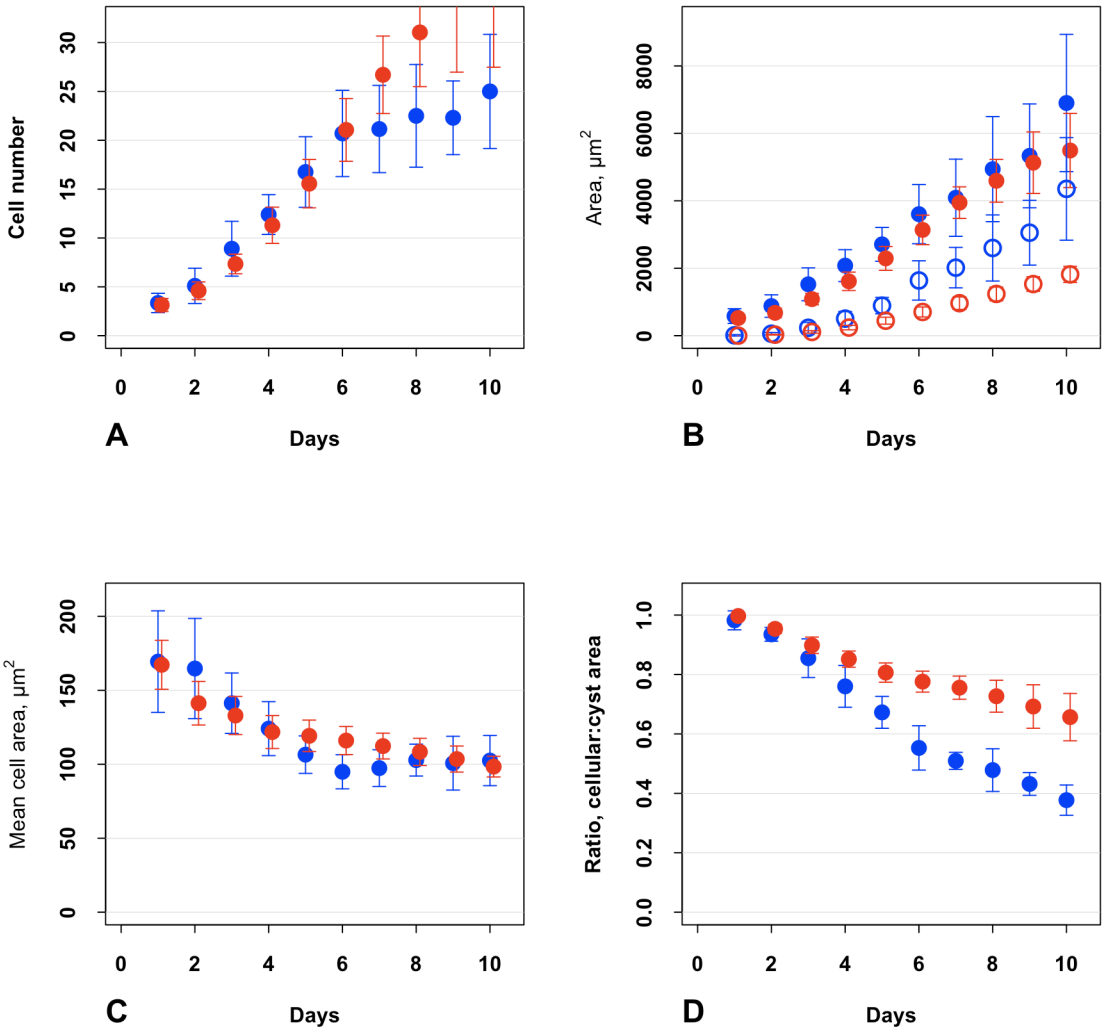


Figure 3.18. CYSTOGENESIS measures when the axis of CELL division is reversed. Experiments followed the same design as in Figure 3.7. Measures, designations, and symbols are also the same. ISMAs used the parameter values in Table 3.3, except for *divisionReg*, which was set to 3.

3.3.9. In silico CYST growth with no LUMINAL CELL DEATH

Cell death contributes to cystogenesis, but it remains unclear to what extent it is essential. In order to explore the consequences of decreased CELL DEATH frequency, we executed simulations in which we reduced *deathRateLumen* from 0.02 to 0.0. We did not

alter the probability of CELL DEATH in CELLS contacting MATRIX. We noted no significant difference in CELL number during the first six DAYS of growth, but during days 7 through 10 mean CELL number was 10-15% higher than observed during control ISMA growth (Figure 3.19A). The observed standard deviations also increased. We observed a smaller percentage of SLSL CYSTS than in control simulations, especially during DAYS 6 to 10 (Figure 3.20B). Values for CYST area, LUMEN area, CELL size, and the ratio of CELLULAR to CYST area were similar to control values (Figure 3.19), while the percentage of single LUMEN CYSTS decreased slightly (Figure 3.20).

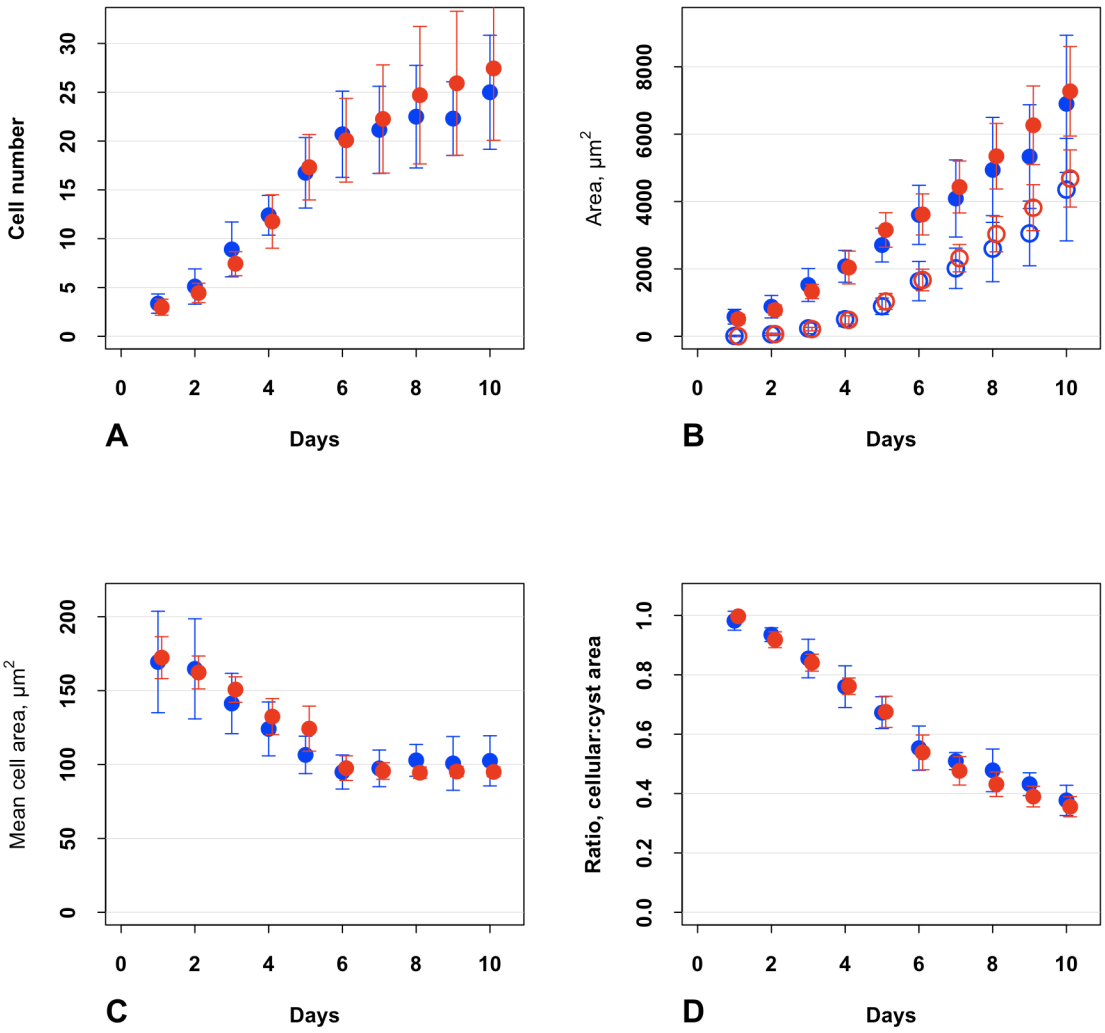
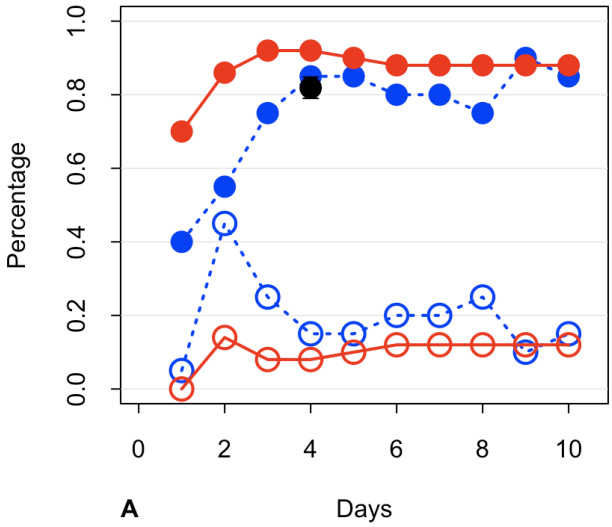


Figure 3.19. CYSTOGENESIS measures with no LUMINAL CELL DEATH.
Experiments followed the same design as in Figure 3.7. Measures, designations, and symbols are also the same. ISMAs used the parameter values in Table 3.3, except for *deathRateLumen*, which was set to 0.



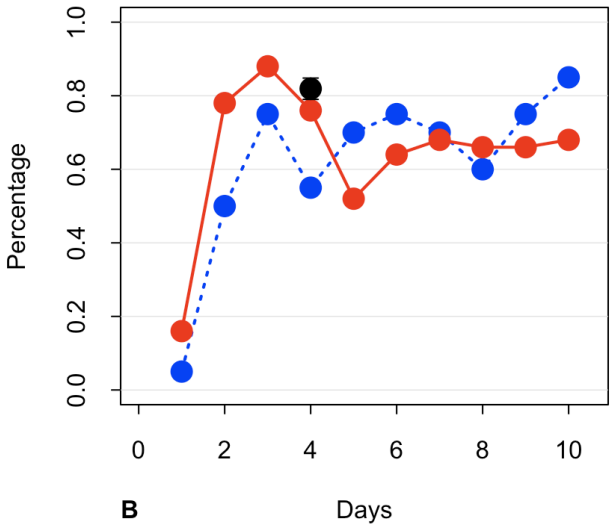


Figure 3.20. Percent of CYSTS with different numbers of LUMENS with no LUMINAL CELL DEATH. The experiments are the same as in Figure 3.19. Designations and symbols are the same as in Figure 3.8.

3.3.10. Simulated CYST growth with delayed CELL POLARIZATION

Delayed cell polarization is believed to contribute to the differences in cyst growth in Matrigel and collagen [51], although it is possible that a lower initial rate of cell clustering and a slower growth rate might be factors as well. To explore the effect of

delayed POLARIZATION on ISMA CYSTOGENESIS, we increased the value of *polarDelay* from 42 (equivalent to 21 hours) to 130 (equivalent to 65 hours). Relative to controls, CELL number increased at an equivalent rate during the first six DAYS, but was larger during DAYS 7–10 (Figure 3.21A). CELL POLARIZATION (data not shown) and LUMEN formation occurred later than in controls (Figure 3.21B). The area taken up by CELLS remained roughly constant, but the delay in LUMEN formation and resulting smaller LUMENS caused the ratio of CELLULAR area to total CYST area to be significantly larger than control values during DAYS 2-8 (Figure 3.22). Not surprisingly, there were fewer single and multiple LUMEN CYSTS during the first three DAYS. When LUMEN formation began, however, it often resulted in multiple LUMENS (> 80% for DAYS 4–6); SLSL CYSTS were observed infrequently. As LUMENS expanded and merged during the later stages of growth, the frequency of SLSL CYSTS increased. The percentage of DYING CELLS not contacting the MATRIX was significantly larger at DAYS 4–10, indicating that many of these CELLS DIED as LUMEN expansion occurred (data not shown). Some of these in silico results reflect those observed within growth in collagen, but it seems unlikely that delayed cell polarization in vitro is solely responsible for those differences.

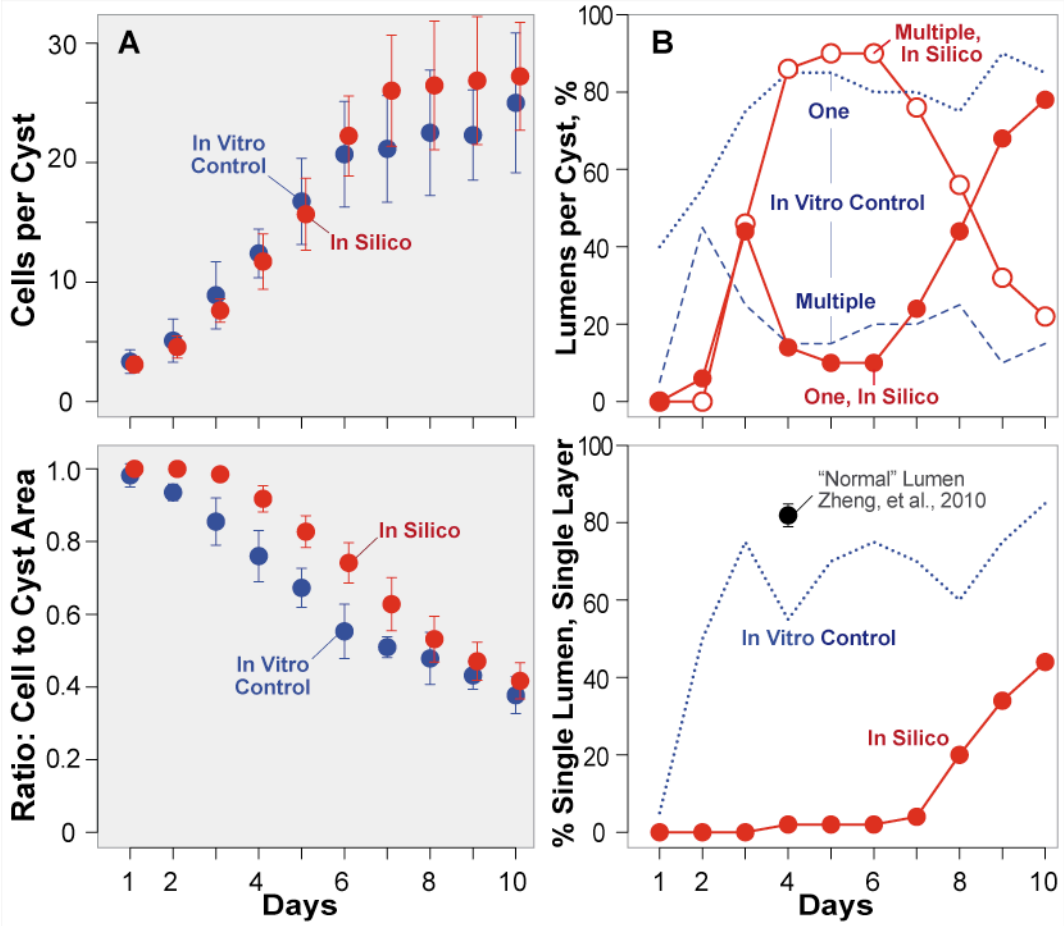


Figure 3.21. CYSTOGENESIS measures and percentages when CELL POLARIZATION was delayed. ISMA simulations executed with the parameters values from Table 3.3 except that CELL POLARIZATION was delayed as described in the text. Left: Mean values and standard deviations for CELL number per CYST (top panel) and ratio of CELLULAR to CYST area (bottom panel). Right: Percentage of CYSTS with single, multiple, and SLSL LUMENS. Designations and symbols are the same as in Figure 3.8.

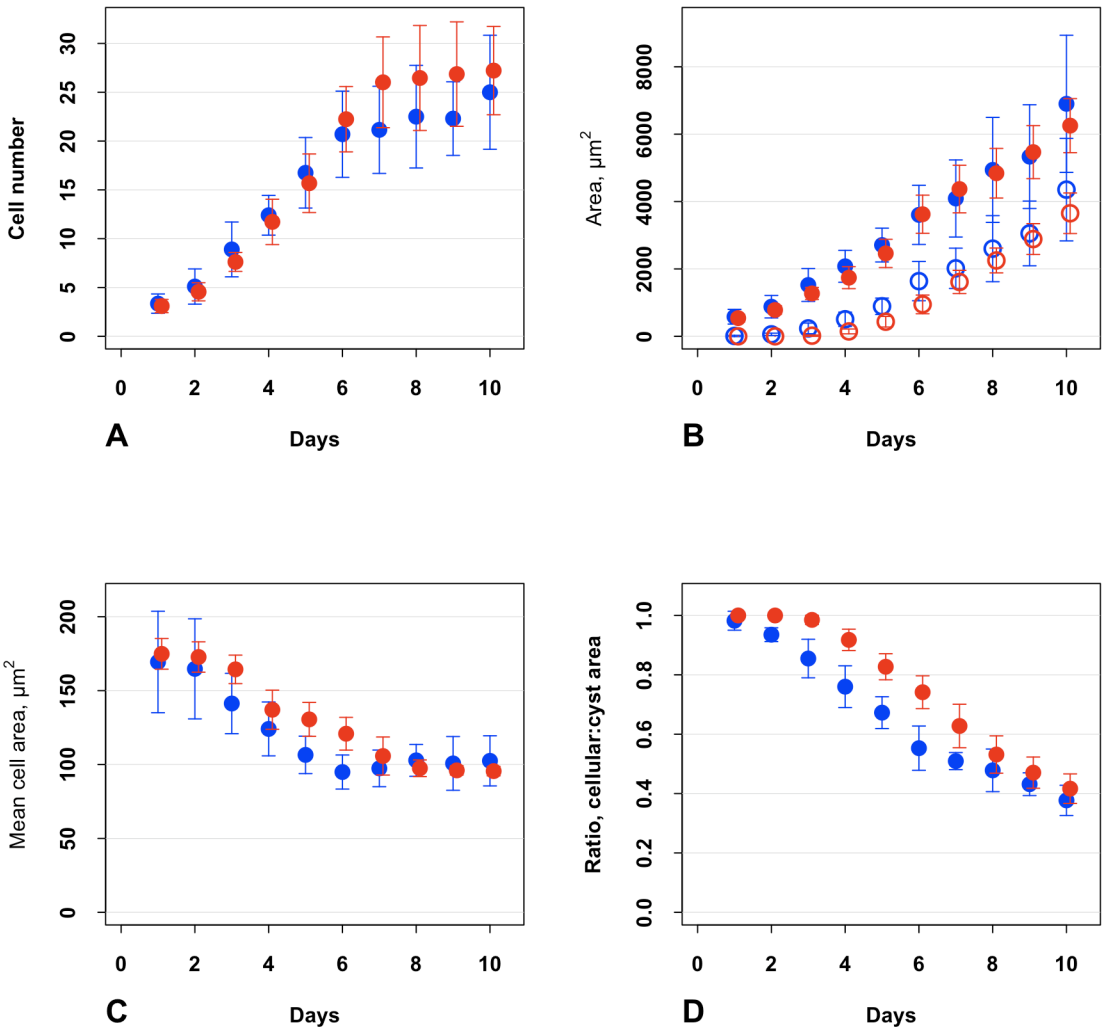


Figure 3.22. CYSTOGENESIS measures when CELL POLARIZATION was delayed.
Experiments followed the same design as in Figure 3.7. Measures, designations, and symbols are also the same. ISMAs used the parameter values in Table 3.3, except for CELL POLARIZATION, which was delayed as described in the text.

3.4. Discussion

3.4.1. In vitro observations

Observations reported herein about in vitro MDCK cystogenesis are consistent with those made previously [47, 48, 51]. There is no evidence of behavioral differences between cells within single and multiple lumen cysts. We could not establish a causative connection between the slowing of cell division and the change in cell size. The evidence

indicates that initial lumen expansion is somewhat isochoric: early lumen expansion is primarily a consequence of cell shrinkage. After an interval of lumen expansion and cell shrinkage lasting about six days, cell behavior changes: cell size stabilizes and cells begin to stretch as the lumen continues to expand (Figure 3.6); cell division slows dramatically; the expanding lumen becomes the primary driver of cyst size; and the variance in both cell area and cyst size increases.

3.4.2. Iterative process

Iteratively constructed ISMAs quantitatively mimicked a targeted set of in vitro data and cell behaviors. Measures of ISMA CYSTOGENESIS matched corresponding measures of MDCK cystogenesis over ten days. The pathways and proteins that play influential roles in cell behavior during MDCK cystogenesis are objects of active research and are increasingly well understood. However, knowledge of how specific cell actions and events are choreographed during cystogenesis is still limited. The latter knowledge is needed to begin establishing causal linkages between molecular level events and systemic phenotype.

Previous analogues [5, 15] used a simple representation of a cell: each CELL occupied a single 2D hexagonal grid space. They were falsified when we added qualitative observations about changes in cell size and shape to our targeted attributes list (Table 3.1). To mimic these newly targeted attributes, we needed CELLS to be more fine-grained. To generate the current ISMA, we began with an in silico analogue that had achieved a degree of validation and then conducted in vitro experiments designed to challenge and possibly falsify it. We then reengineered the in silico system to reflect, explore, and challenge new insight provided by the fresh in vitro data. We engineered

new analogues using the CPM method, which provided several capabilities, including enabling CELL size and shape change. To slow the increase in CELL number after DAY 6, we introduced a stable CELL state.

We envision the above in silico-wet-lab cycle continuing indefinitely. It is straightforward to explore the consequences of in silico mechanistic interventions. If these interventions result in altered system behaviors (predictions), it may suggest new in vitro experiments designed to test them. Examples include the effect of delayed polarization on cyst phenotype, the lack of noticeable changes when cell death is inhibited, and the causal link between lumen size and cell stabilization. Furthermore, we expect a change in cell state (cell stabilization at day 6) to be accompanied by measurable changes in gene expression profiles and biochemical signaling.

3.4.3. Improved analogue

The ISMA illustrated in Figure 3.13 achieved all targeted attributes. It was preceded by two earlier versions. These ISMAs differed in the mechanism used to initiate CELL stabilization. We hypothesized that in vitro cells might use knowledge of their internal geometry to sense their perceived stretch and subsequently stabilize. The geometrical mechanism ISMA (GM ISMA) directly tested this hypothesis; each CELL used measures of its area and geometry to determine when to shift to the stabilized state. To achieve a degree of validation required the use of an axiom specifying that stabilized CELLS would be more likely than POLARIZED CELLS to DIE when not in contact with MATRIX. This axiom was implemented in order to decrease the number of CELLS within the LUMEN and thus increase the number of SLSL CYSTS. The GM ISMA was falsified when targeted SMs for the percentage of single LUMEN, multiple LUMEN, and SLSL

CYSTS were strengthened to those achieved in Figure 3.7, as shown in Figure 3.11 and Figure 3.12. It was falsified because the time at which CELLS stabilized was too variable; some CELLS stabilized early, others much later, resulting in very few SLSL cysts (data not shown).

A second version of the ISMA, called the timed stabilization ISMA (TS ISMA), used an internal clock to signal CELL stabilization, resulting in a uniform stabilization time and reducing the variance in CYST size. The TS ISMA survived falsification (Figure 3.9), providing evidence that stabilization time influences SLSL CYST percentages. The GM ISMA axiom specifying that stabilized CELLS would be more likely than POLARIZED CELLS to DIE when not in contact with MATRIX was not needed. The TS ISMA was capable of generating high percentages of SLSL cysts even without this axiom, and so the axiom was removed in that and subsequent ISMAs.

Although the TS ISMA survived falsification, we were not aware of any in vitro evidence suggesting existence of an equivalent internal clock-based mechanism. If such a mechanism did exist, it might be molecularly equivalent to that of cell polarization. Genes that regulate cellular senescence can suppress the cell cycle, and the sirtuin protein SIRT1 is involved in cellular senescence [69, 70]. It is possible a cell-autonomous timing mechanism could exist that depends on the regulation of SIRT1 and its downstream targets, as detailed below. We hypothesized that a mechanism that used the geometry of the LUMEN instead of the geometry of individual CELLS to signal CELL stabilization might bridge that gap and still produce a low variance in stabilization times. We developed the lumen stabilized ISMA (LS ISMA) described within this report to test that hypothesis and discovered that in addition to surviving falsification (Figure 3.7) it generated stabilization

variance between the GM and TS ISMAs. We can surmise a mapping between the LUMEN-based stabilization mechanism and a functionally equivalent in vitro mechanism in which apical sensory input to each cell provides it with information that correlates to lumen size. Current evidence supports the hypothesis that cells in the cyst wall can sense lumen size. One mechanism by which cells achieve this is by the tension generated at the luminal membrane by membrane stretching. This tensional information is transduced by the subapical F-actin network, which acts both as a scaffold for maintaining luminal integrity, as well as a region for aggregation of recycling endosomes that regulate the protein and lipid composition of the apical plasma membrane. Thus regulators of this F-actin network can regulate lumen and cyst size. Potential molecular mechanisms are detailed below.

We should seek additional, in silico mechanisms that are equally effective in enabling ISMAs to achieve validation targets. Given phenomena, what hypothetical generators (and measures) might generate them? Studying an inverse mapping requires multiple, seemingly plausible hypotheses, which then compete against each other during simulation experiments as done here. After falsification and validation using the IR Protocol, those that survive spawn additional, more refined hypotheses. Furthermore, having multiple mechanistic options for realizing the same behaviors may be biomimetic in that it marginally increases system robustness. An example of a potential additional in silico mechanism is that of time-dependent dynamic parameters, which might assist in the exploration of finer-grained, intracellular molecular behaviors. The current ISMA contains a small number of implicitly dynamic parameters (such as the time that elapses between CELL DIVISION events), which effectively change when CELLS change state, but in

general all parameters are fixed for the duration of the simulation. If in vitro data were targeted that demonstrated, for example, the build-up of certain proteins along the plasma membrane, dynamic variables could be implemented that controlled the amount of that PROTEIN within the analogue.

We note that the standard deviations of quantitative in silico values were generally lower than those observed in vitro (Figure 3.7), though still acceptably similar. The most likely reason for this difference is that although inherently varied, the in silico simulation is computational and follows a set series of steps. If some in vitro cysts developed along different trajectories it might not be observed due to the somewhat small number of cysts observed. This could be addressed through additional in vitro experimentation, in which groups of divergent cysts would be noted. We did note that the TS ISMA exhibited very similar standard deviations for cyst and lumen area (Figure 3.9), suggesting that the close link between CELL number and LUMEN size implied by the LS ISMA may not fully explain the observed growth phenomena.

3.4.4. Potential molecular counterparts to TS and LS ISMA mechanisms

The sirtuin protein SIRT1, a protein deacetylase, has been shown to be involved in cellular senescence, the early phases of which are reversible [70]. This is associated with downregulation of SIRT1 and an increased function of cyclic AMP-regulated kinase (AMPK) [71]. SIRT1 downregulation increases the stability of the AMPK regulator LKB1 and thus inhibits cell proliferation [72]. A possible cell autonomous timing mechanism that could cause cell stabilization might involve downregulation of SIRT1 and upregualtion of LKB1.

Evidence suggests is possible that the tension generated at the luminal membrane

is transduced by the subapical F-actin network. It maintains luminal integrity and allows recycling endosomes to aggregate, regulating the protein and lipid composition of the apical plasma membrane. By regulating this F-actin network, cells can control lumen and cyst size. The Rho family small GTPhase Cdc42 is an ideal candidate for such a regulatory role [73]. Cdc42 is recruited to the apical plasma membrane by the Posphatidylinositol (4,5) bisphosphate (PI(4,5)P2) binding protein Annexin 2, and the loss of Cdc42 disrupts MDCK lumen formation ([74]. The GTPase exchange factor Tuba activates Cdc42 at the apical membrane, allowing it to control apical exocytosis and thus expand and maintain the apical plasma membrane [75]. Cdc42, in concert with PI(4,5)P2, also helps to polymerize actin by regulating N-Wasp and Arp2/3, thus maintaining the subapical F-actin network [76]. Cdc42 also prevents excessive apical constriction by antagonizing the GTPase RhoA through p190RhoGAP [77, 78]. Through the multiple regulatory roles of apical vesicle exocytosis, F-actin scaffold maintenance and regulation of apical constriction, Cdc42 is in a central position to control lumen size.

Cdc42 depletion early in MDCK cystogenesis leads to a loss of central lumen formation, further supporting its central role in this process [74]. Overexpressing a WT Cdc42 in drosophila pupal eye cells increases the apical membrane area [78], possibly reducing luminal tension. This lower tensions and higher surface area allows more fluid to be pumped into the luminal space by transmembrane pumps, increasing lumen size until the apical membrane tension is restored. This molecular mechanism presents one potential explanation for how MDCK cells could sense luminal tension and react accordingly.

3.4.5. Challenging ISMA predictions

ISMAs had already achieved all targeted attribute when the work of Zheng et al. [47] was published. Results from their studies provided an independent challenge to ISMA mechanisms and their robustness. The red simulation results in Figure 3.17 are a consequence of two different simulation interventions: making the CELL axis of DIVISION random (Figure 3.16A) and reversing the CELL axis of DIVISION (rotating it 90°)(Figure 3.16B). These predictions are fully consistent with the in vitro results of Zheng et al. As stated previously they defined a normal cyst as one with actin staining at the apical cell surfaces surrounding a single lumen. Included in that definition are our SLSL CYSTS and CYSTS with a single LUMEN. In Zheng et al., when cell division was randomized, the percentage of cysts with single lumens at day 4 dropped from 81.9% to 21.5%, a different of 60.4%. In ISMA simulations, when *divisionReg* was changed from 1 (ordered DIVISION) to 0 (random DIVISION) the percentage of CYSTS with a single LUMEN dropped from 94% to 46%, a difference of 48%, which is quite similar to the decrease observed in vitro (Figure 3.17A). As seen in Figure 3.16B, when the axis of division was reversed, the percentage of cysts with a single lumen dropped from 81.9% to 11.5%, a difference of 70.4%. Within the ISMA, when *divisionReg* was changed from 1 to 3 (reversed division), the percentage of CYSTS with a single LUMEN dropped from 94% to 14%, a difference of 80%. In addition, the in silico results provide a prediction of in vitro behavior that could be challenged through in vitro experimentation. When DIVISION is reversed within the LS ISMA (Figure 3.18) CELL number continues to increase after day 6, most likely because the numerous small LUMENS do not reach a sufficient size to cause CELL stabilization. In strong contrast, when DIVISION is reversed within the TS ISMA (Appendix 1, Figure 2) CELL number stops increasing at day 5 and remains stable

thereafter. Future experiments of the type conducted by Zheng et al. that quantify cytogenesis over longer intervals would provide evidence supporting one or the other mechanistic hypothesis.

If the results of the experiments of Zheng et al. had falsified the ISMAs, we would have had to look into simulation details during execution to see where and how failure occurred. Because we designed the analogues to make mechanism change straightforward, it would have been simple to discover a new set of mechanisms that achieved all the targeted attributes described herein, as well as the new attributes from the challenge experiments. The results of these experiments demonstrated two important uses for this class of in silico models. 1) They can be used to execute in silico experiments designed to test hypotheses based on the expected consequences of a mechanistic outcome. Our experience has been that aspects of expected outcomes are typically wrong in one or more ways. 2) The results of simulations can be used to stand as hypotheses about outcomes of similarly designed in vitro experiments. The first use exercises our understanding of the networked relations between ISMA mechanisms and phenotype, including systemic behaviors, and that in turn facilitates thinking more deeply about MDCK cell biology. The phenotypes of descendents of current ISMAs will cover increasingly complex behaviors of MDCK cyst cultures. At that stage, it may become standard practice to conduct many exploratory intervention experiments in silico in order to better focus the design of in vitro experiments, marking a fundamental transition from reverse engineering to forward engineering of complex biological systems.

3.4.6. CELL-level and INTRACELLULAR events

A CELL-level event is one that is visible at the current level of resolution. An

event that maps to an intracellular process (referred to as INTRACELLULAR) can occur without causing a visible change; it is below the current level of resolution. Of the events listed in Table 3.7, the two marked (*) only exist within the in silico system and have no specific in vitro counterpart.

Beyond simply modeling cystogenesis, a purpose of this research has been to instantiate an in silico system in which CELLS, MATRIX, and LUMEN have in vitro counterparts, and when executed the ISMA produces a variety of measurable phenomena that quantitatively mimic MDCK cystogenesis. At the systemic level, we have excellent cystogenesis similarity over ten days for multiple measures. Further analogue improvement will, following additional cycles of the IR Protocol, allow INTRACELLULAR events to become concretized and increasingly fine-grained, thus enabling quantitative in silico-to-in vitro mappings at multiple levels.

Table 3.7. CELL and INTRACELLULAR events that can occur within a simulation cycle.

CELL-Level Events	Map to Intracellular Events
CELL state (& color) changes	*MCell* point assignment*
CELL DIVISION	*CellCycle* updating at simulation cycle 1
LUMEN creation	CELL initiates DYING
LUMEN merging through TJ reorganization	DEATH advances; CELL TA decreases
LUMEN expansion through TJ reorganization	Polarity counter (*polarCounter*) begins
LUMEN expansion without TJ reorganization	Decrement *cycleCounter*
Isolated point engulfed*	Decrement *polarCounter*
CELL perimeter (but not TJs) changes	Decrement *shiftCounter* (TS ISMA only)
DYING complete: CELL disappears	Compute MATRIX and LUMEN contact length, A, TA, & TP
MATRIX removal	Compute G for a potential index change
CELLS change shape	

All CELL events produce a visible change within the ISMA visualization. Events that map to intracellular events result in a change within a CELL, but do not produce a visible change within the ISMA. CELL-level events map to equivalent events between in vitro MDCK cells, lumen, and matrix, while INTRACELLULAR events map to events (less well understood) within in vitro MDCK cells. *: This event exists only within the ISMA system and has no specific cystogenesis counterpart.

All specified events were necessary and essential for achieving targeted SMs. For CELL-level events, the mappings are clear: they are direct and quantifiable. INTRACELLULAR events, axioms, and protocols are below the current level of resolution. There is no requirement that a specific INTRACELLULAR event, axiom, or protocol has a cell-level counterpart. We simply hypothesize that the *set* of INTRACELLULAR events, axioms, and protocols—a CELL'S operating principles—has an in vitro counterpart, as illustrated in Figure 1.1. For some INTRACELLULAR events, conceptual mappings are clear. Examples include CELL initiates DYING, DEATH advances, and decrement *polarCounter*. For others, conceptual mappings are less clear. Examples include decrement *shiftCounter* (in the TS ISMA), compute TP, and compute G. The expectation is that, in moving forward, as axioms are replaced by concrete, interacting components (see [7] and the *future experiments* subsection below) clear mappings will be easier to establish and quantify.

A good example of a project in which INTRACELLULAR events are incorporated and to some degree mapped back to those in vitro, is the IBCell model [14, 53]. It is a biomechanical model of MCF-10A cell cystogenesis in which proteins on the outer cell membrane and the extracellular matrix are specifically simulated. The IBCell model successfully mimicked some aspects of cystogenesis, but it remains unclear whether the INTRACELLULAR details are necessary or could be replaced by coarse-grained components. The quantitative data used to validate the model lacked the level of resolution necessary to falsify the intracellular mechanisms.

3.4.7. Cell death and the timing of cell polarization

Surprisingly, CYSTS with little or no CELL death can still be well organized with a

single LUMEN. Reducing CELL DEATH rates (Figure 3.19) altered CYSTOGENESIS details only marginally, primarily because CELL DEATH frequency was already low (Figure 3.14B). Lin et al. [79] hypothesized that apoptosis is crucial for lumen formation in MDCK cysts, but they reached that conclusion based on observations of cystogenesis in collagen culture only. Martín-Belmonte et al. [51] observed that apoptosis within Matrigel cultures is less frequent than within collagen cultures. Within ISMA simulations, earlier LUMEN formation results in more organized CYST growth and fewer CELLS that DIE after losing contact with the MATRIX once LUMENS have formed. It is possible that apoptosis acts simply as a cleanup mechanism within MDCK cysts, but the degree to which it is utilized depends on the environment, the rate of cell growth, and the timing of polarization. Our experiments reducing the rate of CELL DEATH showed that although the rate of CELL DEATH within CYSTS during growth is normally quite low, CELL DEATH still contributes to controlling CELL number and maintaining SLSL CYSTS. It is possible that environmental adjustments may provide conditions in which MDCK cell cystogenesis produces normal SLSL cysts without requiring cell death, as occurs in human alveolar type II epithelial cells [58, 80].

Relative to cystogenesis in Matrigel, cells grown in collagen produce smaller cysts with fewer cells and delayed polarization. That delay might play a role in formation of smaller cysts. However, ISMA experiments showed that delaying POLARIZATION (Figure 3.21) increased cell number and decreased the percentage of CYSTS with single LUMENS. We take those observations as strong evidence that delayed cell polarization alone is insufficient to account for that difference in cystogenesis within collagen and Matrigel cultures.

3.4.8. Future in vitro experiments

As illustrated in Figure 1.1, a goal is to build, expand, and validate in silico mechanistic networks that map to plausible causal linkages between intracellular details and features of MDCK cell phenotype in culture. A prerequisite is to have CELLS capable of achieving increasingly fine-grained and expanding coverage of MDCK cell, cluster, and cyst behaviors under different conditions. Advances in imaging technology have made doing so easier. Similar coverage will be needed of intracellular (subcellular) dynamics, including the behaviors of cell components under different conditions. We anticipate that studies of in vitro MDCK cell cystogenesis using high-resolution, time-lapse microscopy will reveal new behavioral details at each level. Recent studies have employed confocal time-lapse microscopy to understand lumen formation, but only imaged cells for eight hours [81]. Ewald et al. [82] set the standard for long-term time-lapse microscopy in their work on the elongation of mouse mammary ducts, in which they captured individual images every 15 minutes for five days, using high-sensitivity cameras to avoid phototoxicity.

There is ample evidence that tension within the extracellular matrix influences epithelial cell behaviors [83, 84]. Paszek et al. [83] demonstrated that increasing matrix stiffness resulted in tumorigenic behavior in MCF-10A cells. It seems reasonable to expect changes in MDCK cell, cluster, and/or cyst behaviors as Matrigel stiffness, density, and additives are changed. Experiments similar to those within [83], conducted with MDCK cells and for longer durations are needed to expand ISMA coverage of MDCK phenotype in important ways.

Although the underlying in vitro molecular mechanisms to which the TS and LS ISMA map remain unclear, in vitro experiments may indicate one mechanism as being

more plausible. Careful analysis of images generated through time-lapse microscopy is expected to be informative. If the elapsed time between individual cyst polarization and stabilization of division rate or mean cell size are similar between cysts, that would be supportive of the the internal clock mechanism. However, if the interval varied between cysts, that would falsify such a mechanism. If mean lumen size when division rate and cell size have stabilized are similar between cysts, that would support the shift mechanism based on lumen size.

An additional experiment that could validate the LS ISMA would involve disrupting other regulators of luminal tension. Depletion of key regulators of apical tension such as Cdc42 and its partner Protein kinase C zeta/lambda using RNA interference should decrease lumen size. Increasing Cdc42 activity, either by overexpression of WT Cdc42 or a constitutively active form (Cdc42-G12V) would be predicted to increase lumen size. In addition, depletion of RhoA and its effectors ROCK I and ROCK II, which regulate myosin contractility, should also decrease lumen size, since there will be insufficient luminal tension to support lumen expansion. In addition to potentially validating the molecular mechanisms, if disrupting Cdc42 causes cysts to develop multiple lumens and these cysts with smaller lumens fail to stabilize, it will indicate that stabilization is not caused by a timing-dependent mechanism, and is instead caused by a lumen or geometry dependent mechanism.

An ectopic increase in SIRT1 by stable overexpression or knockdown in LKB1 by RNA interference should prevent cysts from stabilizing and promote continued cyst growth. Thus, if cysts grown with cells with overexpression of SIRT1 or LKB1-KD displayed higher cell number at later time points, it would suggest that the molecular

mechanism of cell stabilization was linked to SIRT1 and LKB1.

3.4.9. Future in silico experiments

Five directions for in silico experiments present themselves. The first two require seeking contradictory or supportive literature evidence of in silico experiments. 1) Exploring the consequences of parameter changes will provide insight into ISMA's mechanism-phenotype relationships for which there may be biological counterparts [45]. A full suite of parameter change experiments was conducted using the LS-ISMA; results are presented in Appendix 1. One example is to explore the consequences of changing *deathRateEpi* and *deathRateLumen* (Figure 3.19 and Appendix 1, Figures 61-68), including setting both to 0. Another is to vary *lumenGrowthRate* to explore the effect of increased or decreased LUMEN expansion on in silico CYSTOGENESIS (Appendix 1, Figures 53-60). Addition of any of several compounds to the culture media in vitro will stimulate cyst expansion. Examples include cholera toxin and forskolin. 2) Modify axioms and operating principles to simulate targeted mechanistic interventions. One example (see Results) is to modify the way in which CELLS calculate their axis of DIVISION. Another is to modify how MATRIX is represented in order to explore consequences of altered MATRIX properties on CYSTOGENESIS. Currently, MATRIX is simply a grid space state. Matrix could be represented using a CPM "cell" that offers resistance to CELL advancement. So doing opens the door to exploration of a variety of MATRIX-CELL interactions that can map to proteins altering local matrix properties. 3) Systematically expand the targeted attributes while keeping CELLS atomic. Movies, such as Video S1 from [51] along with the current literature, contain examples of many behaviors beyond the scope of the current ISMAs. Adding any one of the following to the list of targeted attributes will

falsify the current ISMAs. At the cell level: when cells undergo mitosis, they enlarge temporarily and then return to a smaller size; some cells (and cysts) move around during the early stages of cystogenesis; some cells migrate toward each other and cluster together before initiating division; typically, when cells die in contact with matrix, they are flushed into the luminal space where they shrink and disappear. At the cyst level: cysts spin. The process was described by [81] and recently modeled by [85]. Cyst growth may have an additional later stage characterized by significantly slowed expansion, rather than continuing to grow steadily as predicted by the ISMA. The dynamics of lumen merging are more complex than the merging events that occur during simulations. Also, lumens change shape and move within cysts during the initial stages of growth.

4) Increase realism by transforming CELLS from atomic to composite objects. The axioms used by CELLS are placeholders for more fine-grained micromechanisms. The latter can be instantiated in future ISMA descendents. Before we can our turn attention to intracellular processes, we need new ISMAs in which CELLS are composite (and eventually hierarchical) analogues that can achieve essentially the same, targeted SMs as the current ISMAs. Previous reports [7, 57, 58] explained that an in silico analogue (such as the current ISMA) that quantitatively mimics many cell-level phenomena can be used to begin the sequential process of drilling down and establishing plausible, causal linkages between phenotype and molecular level details. Using cross-model validation procedures, the atomic CELL is replaced by a composite CELL where phenomenal axioms are replaced by concrete micromechanisms involving interacting objects that map to subcellular processes and/or components in the referent. 5) Once we have the preceding

composite CELLS, we can expand the list of targeted attributes to include subcellular and intracellular behaviors. Alternatively, expanding the list of targeted attributes can require transforming CELLS from atomic to composite objects. Examples of subcellular and intracellular behaviors include the amount and location of polarization proteins, organelle movement, the organization of the mitotic spindle, formation of a pre-apical patch, location-dependent lipid compartments within the membrane, etc. During cell polarization (as detailed in [67]), PTEN moves to the apical membrane, where it converts PIP3 to PIP2, which binds to Anx2 and assists in the recruitment of Cdc42 to the apical membrane. The task at this stage, while adhering to a strong parsimony guideline, is to add new mechanisms and details that enable validation against the new, targeted attributes, while retaining all of those mechanisms and behaviors that enabled validation during earlier cycles of the IR Protocol. So doing will enable the in silico exploration, falsification, and validation of increasingly complex in vitro MDCK cell behaviors, which will ultimately correlate to in vivo phenotypes of developing epithelial organs.

We hypothesize that the local cause-and-effect relationships (mechanisms) occurring in ISMAs during execution, and thus their morphogenic agenda, have in vitro counterparts. Challenging these alternative hypotheses can be a focus for future in vitro experiments and ISMA refinements.

3.4.10. Conclusion

Through careful application of the IR Protocol, analogues of MDCK cystogenesis in cultures (ISMAs) were developed, falsified, refined, and validated against novel, multi-attribute quantitative data. ISMAs were based on software specifications that enabled in silico behaviors during simulation to achieve degrees of validation: to be

mapped quantitatively to measures of cystogenesis (targeted attributes). Those specifications also enabled hypothesizing that ISMA operating principles, axioms, components, events, and mechanisms have in vitro counterparts. Predictions of substantive mechanistic changes were verified by independent experiments. ISMAs were used to explore and test hypotheses about CELL and CYST dynamics. . The above, coupled in vitro and in silico experiments led to four insights. 1) The axis of CELL DIVISION significantly affects LUMEN number without changing CELL number or CYST size. 2) Reducing the amount of LUMINAL CELL DEATH had limited effect on CYSTOGENESIS. 3) Later stages of cystogenesis, marked by a decrease in the rate of cell division and cessation of the decrease in mean cell size, can be explained by the presence of a new cell state (called stabilized), which differs in a few key behaviors. 4) The same, multi-attribute phenotype can be a consequence of two fundamentally different mechanisms that, in silico, only alter the mechanism of CELL stabilization. By providing a new way of thinking about cystogenesis, ISMA simulations have provided an impetus to explore novel aspects of epithelial morphogenesis.

4. Challenging the ISMA using in vitro collagen data

4.1. Introduction

4.1.1. Differences between collagen and Matrigel in vitro

As noted in [51], MDCK cells exhibit similar qualitative patterns of growth in Matrigel and collagen, with distinct quantitative trajectories. Cells in both culture systems polarize, form expanding lumens, and undergo cell death, but at different rates. Many of these differences can be attributed to the presences of laminin with Matrigel, which causes faster cell polarization and other downstream events. Cysts in collagen have fewer cells, are smaller, develop lumens later, and exhibit more cell death than in Matrigel [51]. In addition, increased cell death is thought critical to lumen formation and expansion. What remains unclear, however, is whether the two systems are the result of cells following a single set of operating principles within different environmental conditions, or if two distinct sets of operating principles are required to generate observed behaviors.

4.1.2. Reasons for undertaking the creation of the ISMA-C

The previously created ISMA, called the ISMA-M within this chapter, successfully survived falsification against quantitative data for MDCK cystogenesis within Matrigel culture. In order to explore how MDCK cystogenesis within Matrigel and collagen differed, additional data was generated that challenged the existing ISMA-M. MDCK cells were embedded in collagen culture and the resulting cysts fixed and imaged as in Matrigel. The resulting quantitative data falsified the existing ISMA-M, necessitating the generation of a new ISMA, called the ISMA-C. This second ISMA

differed only in the parameter settings chosen before conducting experiments—the underlying code was unchanged. Thus, the hypothesis embodied by the ISMA-C is that a single set of operating principles could generate differently parameterized ISMAs that would survive falsification through two distinct data sets.

The process of analyzing the in vitro data, falsifying the existing analogue, and validating a new analogue could lead to new insights about the process of MDCK cystogenesis. The new in vitro data could shed light on the quantitative similarities and differences between MDCK cystogenesis in different culture conditions, including whether cell size ceased to decrease at a similar time as in Matrigel, whether cell number followed a similar trend, and whether the percentage of cysts with multiple lumens would be significantly different. By analyzing whether ISMAs could survive falsification by data from collagen and Matrigel when parameterized differently, we could gain an understanding of the underlying patterns in growth. In addition, analyzing which parameters required changing and which could remain static could offer insight into the specific differences and similarities between the two ISMAs and by proxy their in vitro referents.

4.1.3. Summary of results achieved

The ISMA-C survived falsification through a number of in vitro challenges. The in vitro data was not as robust as that obtained from Matrigel and so the SMs used were less restrictive. Results from the ISMA-C validated for cell number, cyst size, lumen size (during days 5-10), the ratio of cellular to cyst area, and mean cell size. The percentage of CYSTS with single or multiple LUMENS was empirically similar to that observed within the referent in vitro system. In addition CYSTS generated by the ISMA-C showed

increased levels of CELL DEATH, though not as high as that observed within in vitro collagen culture. To create the ISMA-C, parameters influencing CELL number, POLARIZATION time, and CELL DEATH required changing, while those influencing CELL size and LUMEN expansion did not. Based on these results we hypothesize that the mechanisms influencing cell polarization and cell division are different within collagen and Matrigel, but that those influencing lumen expansion and cell size are not. Finally, it seems possible that the differences observed within in vitro cystogenesis in Matrigel and collagen may be the consequences of the same set of operating principles acting within different environmental conditions.

4.2. Methods

4.2.1. In vitro methods

Established protocols for collagen culture of MDCK cyst growth were followed [86, 87]. A single-cell-suspension of MDCK cells was suspended in a solution of buffered, liquefied collagen I (Invitrogen). Three-hundred µl of a 6000 cells/ml suspension was added to each well of an 8-well chambered coverglass-bottom slide (Nunc), resulting in a final cell density of 2500 cells/cm^3, low enough to prevent significant cell clustering. The final concentration of collagen I in the solution was 2 mg/ml, allowing the collagen I to form a gel by incubation in a 37 °C humidified tissue culture incubator before the further addition of 300 µl medium above the gel. One slide of cyst cultures was fixed each day with 4% paraformaldehyde over a ten-day period.

Cysts were stained for immunofluorescence by blocking and permeabilizing the cells with PBS+ containing fish skin gelatin (0.7 w/v) and saponin (0.25% w/v). The plates were next incubated with primary antibodies against GP135 and β-Catenin

overnight at 4 °C. After removal of excess primary antibodies by washing, Alexa-dye conjugated secondary antibodies (Invitrogen) were added and the plates incubated again overnight at 4 °C. Cysts were washed several times with PBS and the samples were stored in PBS containing 0.05% Sodium Azide to prevent microbial growth. Cysts were imaged using a Zeiss 510 laser-scanning confocal microscope (UCSF Helen Diller Family Comprehensive Cancer Center). Images were quantified using empirical image analysis. The number of cells and the areas and perimeters of lumens and cysts were recorded for ten cysts for each day. Isolated cells and cells without either matrix or lumen contact were also counted.

4.2.2. In silico methods

The methods used to create the ISMA-C were to the same as those used to create the ISMA-M. In accordance with the Iterative Refinement Protocol, the existing ISMA was first falsified by the introduction of new quantitative data, in this case from MDCK cystogenesis in collagen. Initial empirical parameter tuning produced an ISMA-C that survived falsification for most quantitative measures, which is described in Appendix 2. This early ISMA-C was falsified when challenged with the observation that cell death is significantly higher in collagen than in Matrigel. Further refinement resulted in the current ISMA-C, which was created through more in-depth parameter tuning and in silico experimentation.

To adjust for the decreased number of cells observed during early cyst development, the value of *clusterProb* was lowered from 0.8 to 0.7. Even with fewer CELLS at DAY 1, the overall CELL number remained too high, and thus *cellCycle* was adjusted down from 100 to 70. Following that, *polarDelay* was tuned so that

POLARIZATION times were analogous to those observed in vitro, which resulted in reasonably similar LUMEN and CYST sizes. At that point CELLS stabilized later than observed because the size of the LUMENS was much smaller than for the ISMA-M, so the value of *stableRatio* was lowered to make CELLS stabilize earlier. The value of *stableCycleDelay* was adjusted to fine-tune the observed CELL number after stabilization. Achieving increased CELL DEATH proved challenging and required adjusting *deathRateLumen*, and *spindleRandom*, as well as fine-tuning other parameter values. Decreasing the value of *cellCycle* caused increased CELL DIVISION but also increased CELL DEATH, for instance. Altered parameters are shown in Table 4.1.

Table 4.1. Parameters changed for simulation of collagen results.

Parameter	Reason parameter changed	Original value	New value
cellCycle	Cells divide less frequently in collagen	70 simulation cycles	100
polarDelay	Cells polarize later in collagen	42 simulation cycles	120
shiftDelay	Reduced to allow CELLS in TS ISMA to stabilize at the observed time	200 simulation cycles	190
clusterProb	Less cell clustering occurs in collagen	0.8	0.7
dyingShrinkRate	More dying cells are observed within collagen	9 grid points	4
stableRatio	Lumens in collagen are smaller	0.5 grid points	0.35
stableCycleDelay	Cell division does not decrease as much after stabilization in collagen	0.85	0.8
deathRateLumen	More apoptotic cells without matrix contact are observed in collagen	0.02	0.03
spindleRandom	More apoptotic cells without matrix contact are observed in collagen after stabilization	0	40%

Parameters that were changed in order to allow the ISMA-C to mimic growth in collagen, along with the in vitro observation that justified that change, and the before and after values. For the LS ISMA *shiftDelay* and *stableRatio* were set to 140,000 and 0.35, while for the TS ISMA *shiftDelay* and *stableRatio* were set to 190 and 1000. All units are relational (e.g., grid points instead of μM, simulation cycles instead of hours).

Because fewer cysts were analyzed in collagen culture than in Matrigel (10 versus 20), the data represented a less accurate sample of quantitative measures of MDCK cystogenesis. As a result, the validation targets were relaxed, especially for lumen area. Accordingly, to survive falsification via SM1, ≥ 50% of simulations must achieve the validation target (±25% of the in vitro mean value) for 8 out of 10 days, or for 5 of the final 6 days for lumen area. In order for an ISMA to surfive falsification when evaluated using SM2, the value of SM2 must be ≤ 0.25 for 7 of 10 days.

In addition to executing experiments that demonstrated that the ISMA-C is capable of surviving falsification when challenged with in vitro data from MDCK cystogenesis in collagen, other experiments were undertaken. In order to evaluate whether a timed shift ISMA would generate similar results when parameterized to mimic cystogenesis in collagen as when parameterized to mimic cystogenesis in Matrigel, a TS ISMA-C was created. This was achieved through raising *stableRatio* to 1000 to prevent LUMEN size-based stabilization and then tuning the value of *shiftDelay* until the ISMA survived falsification. Additional experiments were conducted with the ISMA-C that followed a similar protocol as the ISMA-M. Notably, the axis of DIVISION was randomized or reversed and LUMINAL CELL DEATH was eliminated. The large-scale parameter sweeping performed on the ISMA-M was not judged necessary in this case, because the additional knowledge gained for the ISMA-M was not significant, and because during the process of empirically tuning ISMA-C parameters to achieve similarity, many of these settings were explored. It would be expected that parameter sweeping would result in similar results for the ISMA-C and ISMA-M.

4.3. Results

4.3.1. Quantitative results in vitro

To understand the differences between cyst development within Matrigel and collagen culture, MDCK cells were grown and observed in 3D collagen culture for one to ten days and analyzed quantitatively each day. Cysts developed consistent with previous observations [51]. Cells suspended in collagen formed small clusters during the first 24 hours, and cells polarized during days 2-4. Lumens formed later than within Matrigel, but all cysts had lumens by the end of day 4. The deviation from an ideal circle ranged between 3 and 8% at day 3, slightly higher than within Matrigel.

As with Matrigel, we measured and recorded cyst and lumen area and perimeter, cell number, the number of single and multiple lumen cysts, and single lumen single (cell) layer cysts (SLSL). The results are graphed in Figure 4.1 and Figure 4.2. From these results we calculated mean cell area and the ratio of total cellular area to total cyst area.

Cell number increased exponentially through day 6, and then leveled off and increased at a constant rate. The variance in cell number per cyst differed compared to Matrigel, perhaps because of the smaller number of cysts measured (Figure 4.1A). Cyst and lumen area increased monotonically (Figure 4.1B), except during day 6. Mean cell size decreased unevenly through day 6 (Figure 4.1C) and then leveled off at roughly the same time that cell division slowed. Cell size variance decreased sharply after day 5. As in Matrigel, we did not find a strong correlation between mean cell size and other cyst measurements, including cell number, lumen size, lumen number, or lumen perimeter/cell number. The ratio of total cellular area to cyst area (Figure 4.1D) indicated that the portion of cyst occupied by cells decreased as cysts expanded (and thus the portion

occupied by lumen increased). As in Matrigel the ratio decreased quite steeply between days 5 and 6 with even less overlap; almost all cysts at day 5 had a ratio higher than 0.75 and almost all cysts at day 6 had a ratio lower than 0.7. These observations taken together indicated a shift in cell behavior occurred at approximately day 6 in collagen, similar to the timing of the shift observed in Matrigel, but slightly later.

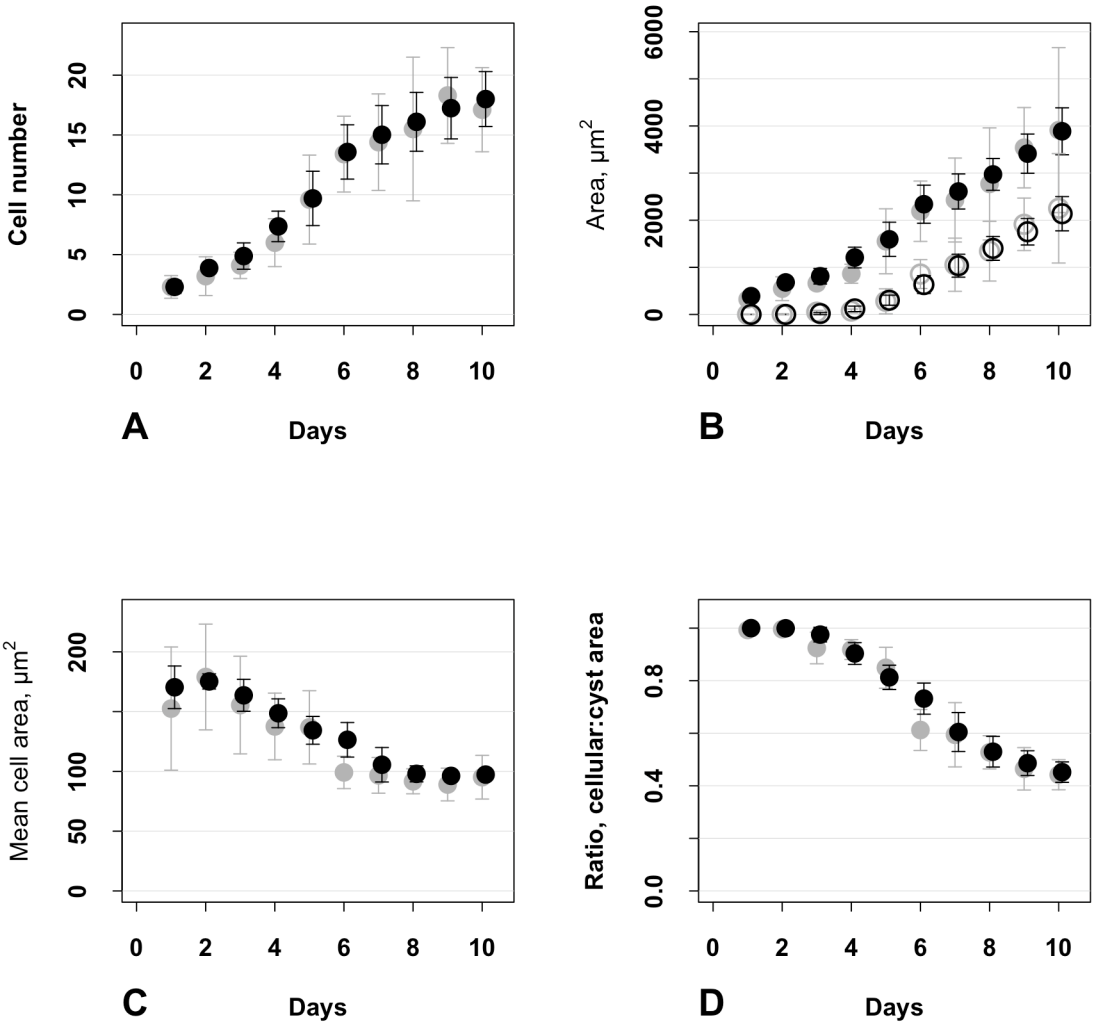


Figure 4.1. Quantitative values for in vitro and in silico cystogenesis in collagen.
Mean values and standard deviations for (A) cell number per cyst, (B) cyst and lumen area, (C) mean individual cell area and (D) ratio: cellular to cyst area. Gray: in vitro data taken each day for ten days from 10 cysts. Black: data taken from 50 CYSTS over ten DAYS using specified parameter settings.

4.3.2. Lumen percentages in vitro

During the first two days of growth in collagen some cysts developed lumens, while others had no visible lumen, but all cysts had lumens by day four. Multiple lumens appeared in a number of cysts, with the percentage fluctuating between 10% and 50% from days 4 to 10 (Figure 4.2A). After day 3, the percentage of SLSL cysts ranged between 50% and 80% (Figure 4.2B), similar to that observed in Matrigel. As in Matrigel, in cases where single-lumen cysts did not have a single layer of cells, usually only one or two cells did not contact the lumen or extracellular matrix. Unlike in Matrigel, the percentage of cysts with multiple lumens did not decrease over time, but the small sample size (ten cysts at each time point) makes generalizations difficult.

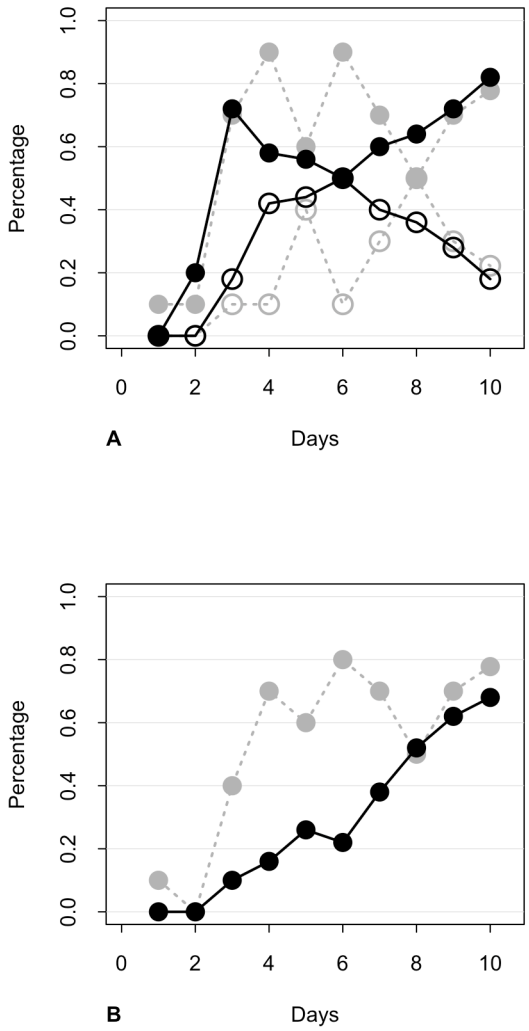


Figure 4.2. Percentage of cysts with specified types of lumens.
(A) Percentage of cysts with single (solid circle) or multiple (open circle) lumens. (B) Percentage of cyst with a single lumen surrounded by a single layer (SLSL) of cells. Gray: in vitro data for 10 cysts taken each day for ten days. Black: in silico data for 50 CYSTS using specified parameter settings. Solid lines: continuous ISMA-C CYST growth. Dotted lines: discrete MDCK cyst growth in collagen.

4.3.3. Quantitative results in silico

ISMA-C CYSTS were structurally similar to cysts grown within collagen. CYSTS began with 1-3 CELLS at DAY 0. CELLS POLARIZED and formed LUMENS within the first three DAYS (Figure 4.2). LUMENS and CYSTS expanded at a rate similar to that observed in vitro. In some CYSTS, a single LUMEN surrounded by a single layer of POLARIZED

CELLS formed within the first four DAYS (Figure 4.2). In other CYSTS, multiple LUMENS formed, each separated by an independent layer of CELLS, such that no CELL contacted more than one LUMEN.

As in collagen, ISMA-C CELL number also exhibited two growth phases, with the rate of CELL DIVISION decreasing after day 6 (Figure 4.1). LUMEN and CYST size increased at rates similar to those observed in vitro, but standard deviations were smaller. CELL size also decreased at a rate comparable to in vitro, and its standard deviations were also smaller during the first five days of growth. The values of Similarity Measure 1 and 2, as well as SSM1 are shown in Table 4.2, Table 4.3, and Table 4.4. As indicated by the values of SM1 and SM2, ISMA-Cs produced quantitative results similar to in vitro values. ISMAs were executed using the parameter settings in Table 4.1 where changed, and Table 3.3 where unchanged. CYST and LUMEN area were scaled by 2.25 μm^2 and perimeter by 0.75 μm as in the ISMA-M.

4.3.4. Lumen percentages in silico

Simulations using ISMA-Cs produced single and multiple LUMEN CYSTS at frequencies comparable to those observed in vitro, (Figure 4.2A), though the rate of multiple lumen cysts was somewhat higher. The percentage of SLSL CYSTS (Figure 4.2B) was lower than observed in vitro, though it increased steadily over time as LUMENS merged.

4.3.5. Similarity measures

The ISMA-C achieved the required SM1 values for cell number, cyst and lumen area, mean cell area, and the ratio of cellular to cyst area. The values of SSM1 are included in Table 4.3, and illustrate the high level of variance within the in vitro results.

ISMA-Cs achieved similarity for SM2 measure for cell number, cyst size, mean cell area, and the ratio of cellular to cyst area. Due to the high variance of luminal size within collagen culture, the ISMA-C did not achieve SM2 for lumen size.

Table 4.2. SM1 values for the ISMA-C.

Day	Cell number	Cyst area	Lumen area	Mean cell area	Ratio: cellular to cyst area
1	0.74	0.72	0.0	0.98	1.0
2	0.0	0.1	0.0	1.0	1.0
3	0.7	0.54	0.18	1.0	1.0
4	0.25	0.1	0.2	0.98	1.0
5	0.68	0.76	0.38	1.0	1.0
6	0.9	0.82	0.5	0.38	0.74
7	0.84	0.9	0.72	0.86	0.94
8	0.88	0.96	0.78	0.96	0.94
9	0.92	0.94	0.86	0.98	0.94
10	0.96	0.96	0.9	1.0	0.96

Values in black did not achieve the validation target.

Table 4.3. SSM1 values for MDCK cystogenesis in collagen.

Day	Cell number	Cyst area	Lumen area	Mean cell area	Ratio: cellular to cyst area
1	0.4	0.6	0.0	0.7	1.0
2	0.2	0.7	0.0	0.8	1.0
3	0.5	0.9	0.3	0.5	1.0
4	0.3	0.7	0.4	0.8	1.0
5	0.6	0.5	0.4	0.8	1.0
6	0.6	0.6	0.5	0.9	0.9
7	0.7	0.6	0.4	0.9	0.9
8	0.4	0.2	0.3	0.9	1.0
9	0.9	0.8	0.6	0.9	0.9
10	0.89	0.22	0.22	0.89	1.0

Values in black would not have achieved the validation target if they were from simulation experiments.

Table 4.4. SM2 values for MDCK cystogenesis in collagen.

Day	Cell number	Cyst area	Lumen area	Mean cell area	Ratio: cellular to cyst area
1	0.18	0.04	0	0.23	0.02
2	0.38	0.33	1.14	0.21	0.01
3	0.04	0.01	0.36	0.18	0.04
4	0.16	0.05	0.08	0.12	0.0
5	0.15	0.22	0.6	0.13	0.04
6	0.07	0.12	0.06	0.02	0.05
7	0.12	0.22	0.30	0.02	0.08
8	0.23	0.31	0.29	0.05	0.01
9	0.07	0.12	0.13	0.1	0.08
10	0.08	0.32	0.35	0.17	0.04

Values in black did not achieve the validation target.

4.3.6. TS ISMA in collagen

To further explore the mechanisms behind CELL stabilization and the differences in MDCK growth in collagen and Matrigel, we changed the parameters *shiftDelay* (from 140,000 to 190) and *stableRatio* (from 0.35 to 1000) to create a TS ISMA-C. The quantitative results of the TS ISMA-C were very similar to those generated by the LS ISMA-C (Figure 4.3 and Figure 4.4). The value of *shiftDelay* used to generate these results was found through empirical tuning, and we noted that the best value (190) for the TS ISMA-C was quite close to the best value for the TS ISMA-M (200), resulting in a mean CELL stabilization time of 6.6 DAYS for the ISMA-C versus 5.25 DAYS in the TS ISMA-M.

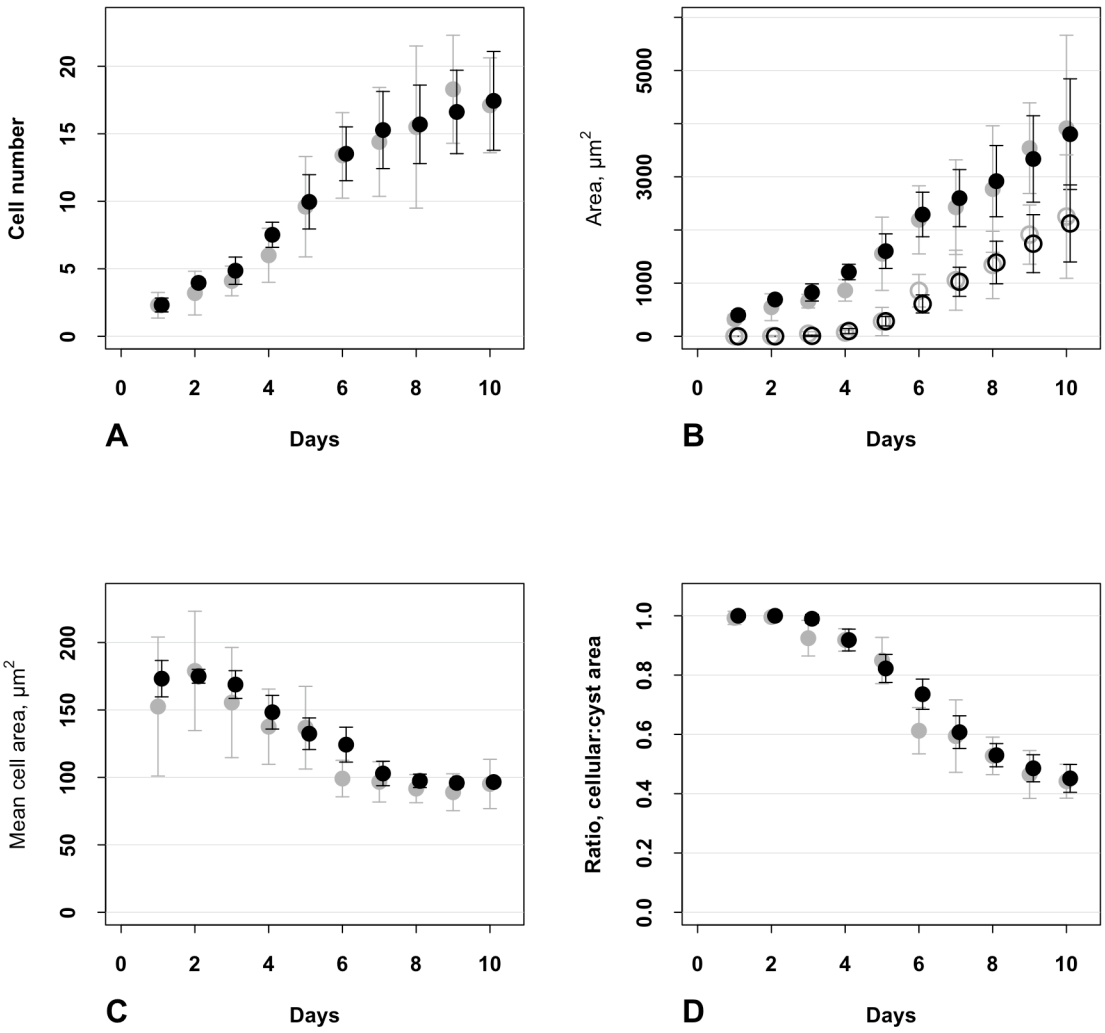
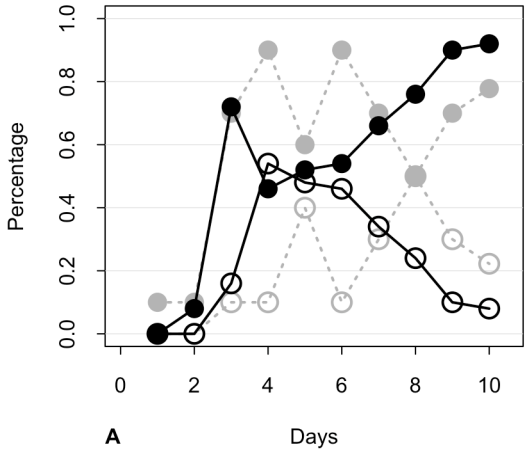


Figure 4.3. CYSTOGENESIS measures for TS ISMA-C.
Experiments followed the same experimental design as described in the text. Measures (black) were taken during CYSTOGENESIS. In vitro data are provided (gray) for comparison**.** Designations and symbols are the same as in Figure 3.7. TS ISMA used the parameter values in Table 3.3 and Table 4.1 except for *stableRatio*, which was set to 1000 and *shiftDelay*, which was set to 190.



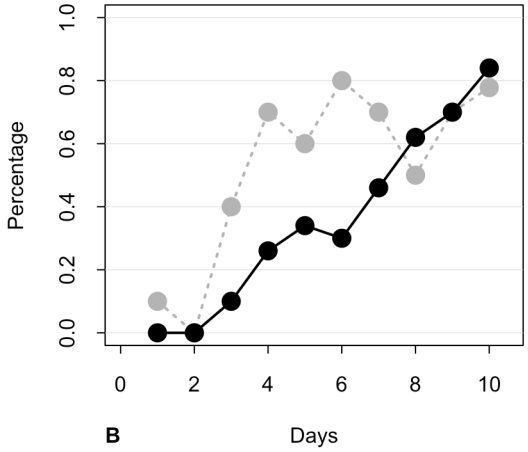


Figure 4.4. Percent of CYSTS with different numbers of lumens for TS ISMA.
The experiments are the same as in Figure 4.3. Designations and symbols are the same as in Figure 3.8.

4.3.7. Cell death

Cell death is commonly observed within the lumen of MDCK cysts during cystogenesis in collagen [51]. One of the challenges when developing the ISMA-C was to increase the amount of cell death observed in order to mimic that observed in vitro. CELL DEATH did occur during ISMA-C execution, but at somewhat lower frequencies than observed in vitro (Figure 4.5). As noted in Chapter 4 Methods, we increased the

average duration between a CELL initiating DEATH and disappearing to be 19 simulation cycles, equivalent to 9.5 HOURS. This was considered an acceptable interference because the duration of apoptosis within MDCK cells in vitro has not been quantitatively established. Although the ISMA-C did not fully mimic the level of cell death observed within MDCK cystogenesis in collagen, it did exhibit higher CELL DEATH than the ISMA-M.

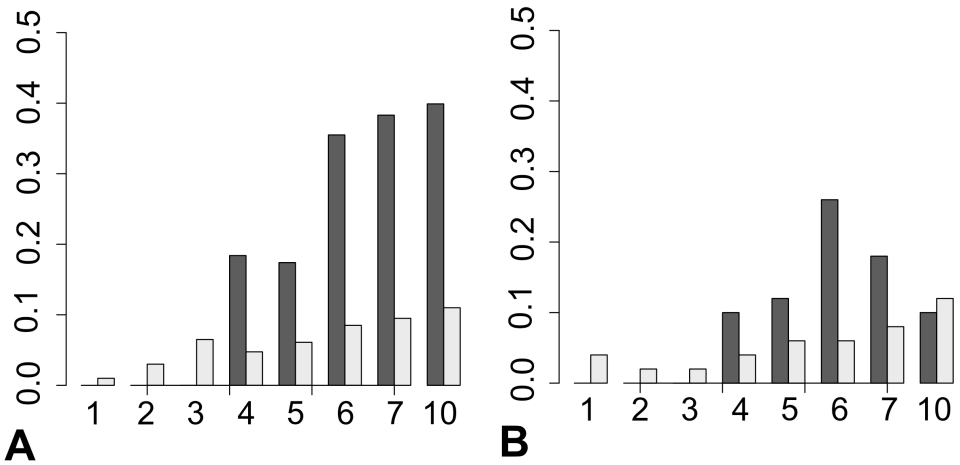


Figure 4.5. Percentage of cysts with dying cells in collagen.
(A) In vitro data reproduced from [51]. (B) ISMA data from 50 CYSTS over ten DAYS. Dark gray bars: percentage of cysts observed to have apoptotic cells without matrix contact. Light gray bars: percentage of cysts observed to have apoptotic cells with matrix contact.

4.3.8. Altered CELL DIVISION orientation in silico

In Chapter 3 we discussed the results of Zheng et al. [47], who measured the consequences of disrupting cell division orientation on MDCK cyst morphology. We executed similar experiments using both the ISMA-M (see Chapter 3) and the ISMA-C. We first altered CELL DIVISION so that all CELLS divided with a random orientation, though it is worth noting that the ISMA-C has a default *spindleRandom* value of 40 rather than 0, meaning that the increase in randomness was less compared to the equivalent experiment with the ISMA-M. The results (Figure 4.6) were similar to those observed with the ISMA-M, although the percentage of single LUMEN and SLSL CYSTS was lower

for the ISMA-C. The altered ISMA-C produced less than 10% SLSL CYSTS and more than 30% multi-LUMEN CYSTS at DAYS 3 through 9. The quantitative aspects of CYST growth were not significantly altered.

In a second experiment, Zheng et al. targeted LGN to the apical membrane. So doing rotated the axis of division by 90°, thus reversing cell division orientation. We conducted a similar experiment on the ISMA-C by modifying CELL DIVISION so that the axis of DIVISION was parallel, rather than perpendicular to the LUMEN edge. During this experiment *spindleRandom* was set to 0. The results were very similar to those produced by the ISMA-M, with many multi-LUMEN and very few SLSL cysts. Randomizing or reversing the axis of CELL DIVISION for the ISMA-C decreased the percentage of single LUMEN and SLSL CYSTS in a similar fashion as for the ISMA-M (Figure 4.6).

The result of altering CELL DIVISION within the ISMA-C provides an additional prediction about cell division and growth within MDCK cystogenesis in vitro. Based on these results, we would expect that LGN-KD and LGN-apical cells grown in collagen would exhibit similar behaviors as those grown in Matrigel, though with a lower percentage of cysts with single lumens.

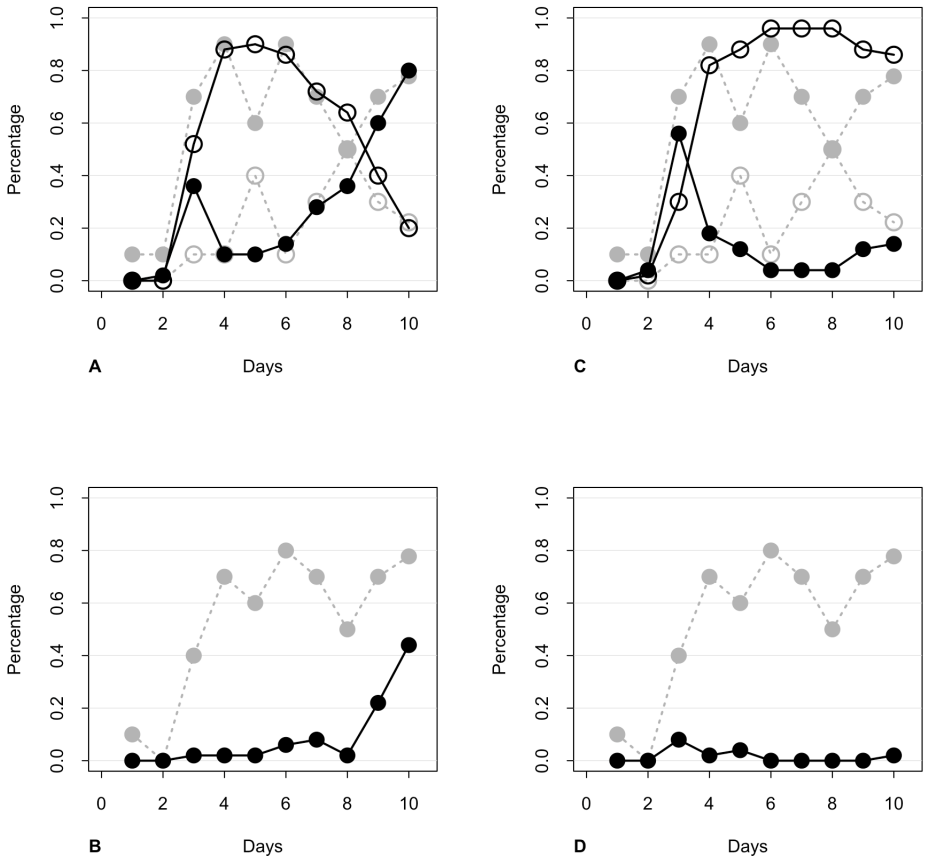


Figure 4.6. Percentage of ISMA-C CYSTS with varied LUMEN number when the axis of CELL DIVISION is abnormal.
Top panels: the percentages of CYSTS that have single (solid black circles) or multiple (open black circles) LUMENS when the axis DIVISION is (A) random or (C) reversed (rotated 90°). The dotted and dashed lines are the corresponding control in vitro values from Figure 4.4. Bottom panels: the percentage of CYSTS that are SLSL (solid black circles) when the axis of CELL DIVISION is (B) random or (D) reversed. The dotted lines are the corresponding control in vitro values from Figure 4.2.

4.3.9. In silico CYST growth with no LUMINAL CELL DEATH

Cell death contributes to cystogenesis, but simulations of cystogenesis using the ISMA-M indicate that it is not required for cystogenesis in simulated Matrigel culture. In order to explore the consequences of decreased CELL DEATH frequency in simulated collagen culture, we executed ISMA-C simulations in which we reduced *deathRateLumen* from 0.03 to 0.0. We did not alter the probability of CELL DEATH in CELLS contacting MATRIX. We noted no significant difference in CELL number during the

first six DAYS of growth, but thereafter the number of CELLS per CYST increased much more than in vitro (Figure 4.7), as did CYST size. The observed standard deviations also increased. We observed a smaller percentage of single LUMEN and SLSL CYSTS than in control ISMA-C simulations, especially during DAYS 6 to 10 (Figure 4.8). Values for LUMEN area, CELL size, and the ratio of CELLULAR to CYST area were similar to control values (Figure 4.7). Reducing LUMINAL CELL DEATH to 0 had a significant effect on the growth of many ISMA-C cysts, more so than for the ISMA-M. This lends support to the existing hypothesis that cell death plays an important role in the development and maintenance of lumens during cystogenesis within collagen culture.

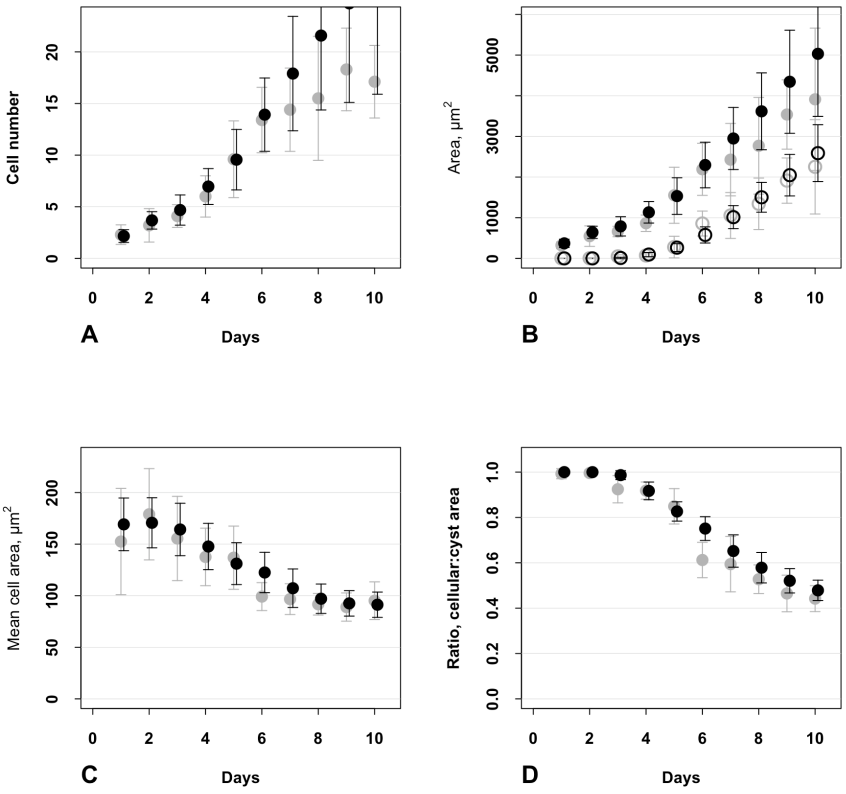


Figure 4.7. CYSTOGENESIS measures with no LUMINAL CELL DEATH.
ISMA simulations executed with the parameter values from Table 4.1 and Table 3.3, except that LUMINAL CELL DEATH was not allowed. Measures (black) were taken during CYSTOGENESIS. In vitro data are provided (gray) comparison**.** Designations and symbols are the same as in Figure 3.7.

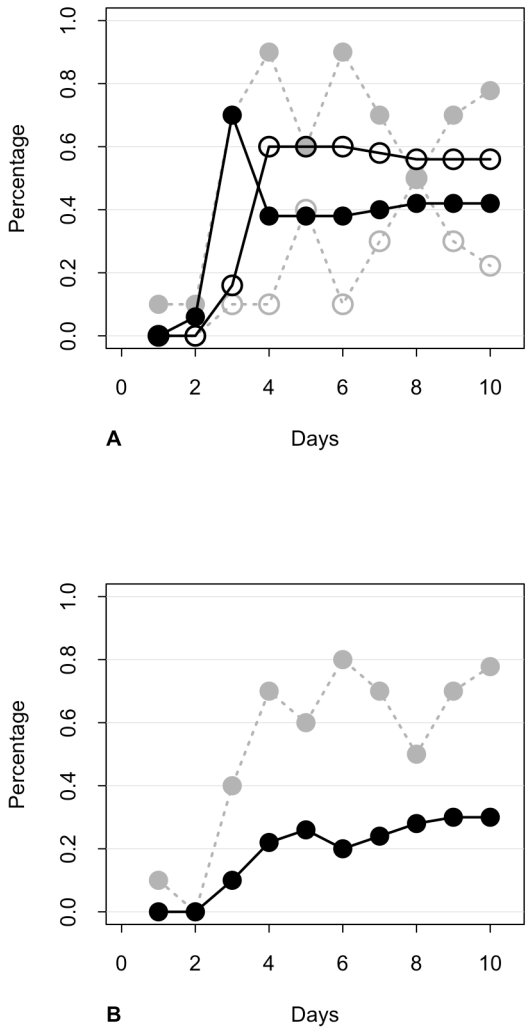


Figure 4.8. Percentage of CYSTS with different numbers of lumens with no LUMINAL CELL DEATH. (A) Percentage of cysts with single or multiple lumens. (B) Percentage of SLSL cysts. Black: in silico data from CYSTS with no LUMINAL CELL DEATH. Gray: in vitro control data.

4.4. Discussion

4.4.1. In vitro observations

The in vitro data presented in this report demonstrates that MDCK cystogenesis in Matrigel and collagen is qualitatively similar and follows similar trends, but differs quantitatively. Cysts form using the same basic cellular behaviors and exhibit a similar trajectory of events: cell division followed by polarization, lumen creation, and cyst and lumen expansion. In both systems single and multiple lumens are present, and cell death is observed in cells with and without matrix contact. The percentages of cysts with single and multiple lumens are quite similar over time. Finally, the difference in lumen size may primarily be the result of later lumen initiation time. Cells in collagen appear to stabilize slightly later than in Matrigel, coinciding with a cessation of mean cell size decrease. It is clear that cells grown in collagen polarize and form lumens later, contain fewer cells, have smaller cyst and lumen sizes, and that more cell death is observed [51]. What is not as readily apparent, but was revealed during quantitative analysis during this project, is that MDCK cysts grown in collagen begin with fewer cells that divide slower. It seems possible that these similarities are the result of MDCK cells following the same operating principles, but in different environmental conditions.

4.4.2. Iterative process

The process of refining the ISMA to mimic MDCK cystogenesis in collagen is an example of the Iterative Refinement Process. After creating an analogue that could survive falsification by qualitative and quantitative data from MDCK cysts grown in Matrigel, we challenged that ISMA with similar data from MDCK cysts grown in collagen. The parameterized ISMA-M did not survive the challenge and was falsified by

these new data. It remained unclear whether the ISMA-M could be reparameterized to survive this challenge, or whether new operating principles would be required to create the ISMA-C. We hypothesized that it would be possible to create an ISMA-C that survived falsification from the new challenge data simply by changing certain key parameters within the ISMA-M. The information gained by analyzing which parameters required changing and which could be left intact would provide useful information about how the two ISMAs differed, and by proxy how MDCK cystogenesis is different within Matrigel and collagen culture.

The first version of the ISMA-C, explained in detail in Appendix 2, achieved a degree of similarity when compared to the quantitative results generated by MDCK cystogenesis in collagen. This early ISMA-C was somewhat similar in cell number, cyst and lumen size, mean cell size, and the ratio of cellular to cyst area. In addition, the percentages of single and multiple lumen cysts appeared similar to those in vitro. However, this ISMA-C displayed very little CELL DEATH, which caused its falsification when qualitatively challenged by the observation that during MDCK cystogenesis in collagen, significantly more cell death is observed than in Matrigel [51]. As a result, the current ISMA-C was developed, which required additional changes to ISMA parameters, as shown in Table 4.1. This version of the ISMA-C did not match the percentages of cell death observed in vitro, but it did demonstrate that it is possible to create an ISMA that has a significant amount of cell death later in growth and still survives qualitative and quantitative falsification.

The current ISMA-C develops fewer SLSL cysts than observed within in vitro culture (Figure 4.2). We believe this is due to the inexact nature of the CPM architecture.

Because the LUMEN develops later the likelihood of CELLS not contacting LUMEN or MATRIX increases. It may also indicate that the current specification does not fully describe the underlying biology. As described below, further refinement of the ISMA-C should produce an analogue that better survives the challenge presented by in vitro data from MDCK cystogenesis in collagen. That refinement, whether it is restricted to parameter tuning or requires modification of ISMA mechanisms, will inform us further of the similarities and differences of the two analogues, and also of the in vitro systems.

4.4.3. Parameters changed from the ISMA-M to the ISMA-C

In order to assess the differences between the ISMA-M and the ISMA-C, we analyzed which parameters were changed and which were allowed to remain the same. As stated in Chapter 4 Methods, it was clear from the results of the quantitative analysis of in vitro growth that both initial cell number and the rate of cell division are different. We found that changing only one of *clusterProb* or *cellCycle* would not allow the ISMA-C to validate, but changing both would. We can conclude that the lower value of *clusterProb* reflects the lower number of initial cells observed in collagen cystogenesis, and that the lower value of *cellCycle* implies that cells in collagen may divide more slowly than in Matrigel. Because cells are suspended within 3D collagen culture, as opposed to within Matrigel where they grow on a single layer, clustering is less likely, thus it follows that *clusterProb* would be lower within the ISMA-C. It is unclear why the rate of cell division seems to be lower within collagen. It is important to note that the rate of cell death observed in the two culture systems during the first three days of cystogenesis is similar, indicating that it is unlikely that cells divide at a similar rate in both systems, since the only factors that control cell number are cell division and cell

death. If cells divided at the same rate and died at the same rate, then only initial cell number could generate the observed difference in cell number, and mathematical modeling demonstrated this was not the case (results not shown). The results of parameter tuning supported this conclusion; we were unable produce an ISMA-C that validated for cell death and cell number during the first three days of growth without changing the value of *cellCycle*.

As observed in [51], cell polarization is significantly delayed within collagen culture, likely due to the lack of laminin within collagen culture. To produce an ISMA-C that survived falsification by this observation, we increased the value of *polarDelay* significantly. We could not create an ISMA that survived falsification against these observations of delayed cell polarization and lumen initiation without changing *polarDelay*. This delayed POLARIZATION resulted in smaller LUMENS and necessitated reducing the value of *stableRatio* to prevent CELLS from stabilizing much later during growth. Because *stableRatio* does not directly map to a known biological component this observation does not falsify the ISMA-C. We did note that although the mean value of in vitro lumen size observed at day 6 was much lower in collagen than in Matrigel (857 μm^2 versus 1638 μm^2), the value of the ratio of cellular to cyst area (0.6 versus 0.55) was similar. This observation indicates that while a lumen-size based mechanism of cell stabilization required changing *stableRatio*, a sensing mechanism based on the value of the ratio of cellular to cyst area might potentially validate without changing that parameter setting.

It was particularly challenging to validate against the rate of cell death for cells not in contact with the matrix, which was significantly higher in collagen culture [51].

Simply increasing the value of *polarDelay* did have an effect, because more CELLS did not have contact with the MATRIX after LUMEN initiation and DIED as a result. However, this effect was primarily noticed during days 5 and 6. A key improvement was attained by lowering the value of *dyingShrinkRate*, which caused DYING CELLS to survive longer. The value of *dyingShrinkRate* corresponds to the amount of time before apoptotic cells in vitro disappear, and as stated in Chapter 3, this is not clearly established. This difference was first noticed during the development of the ISMA-M (Figure 3.15) and it is interesting to note that the rate of CELL DEATH was the only significant observed difference when *dyingShrinkRate* was decreased. We found that increasing *deathRateLumen* did cause the rate of CELL DEATH to increase, but at values higher than 0.03 fewer CELLS were observed to DIE during DAYS 6-10. This result was surprising, but upon consideration does make sense, as if more CELLS DIE earlier, they will not be able to produce new CELLS later, thus the number of CELLS that can DIE at later time points will be lower.

We also changed the value of *spindleRandom* in order to produce CYSTS with a higher rate of CELL DEATH during days 6-10. When *spindleRandom* was set to 0 as in the ISMA-M, very few CELLS lost MATRIX contact, and thus the percentage of CELLS observed to DIE without MATRIX contact was very small. It is quite likely that having some degree of random CELL DIVISION more closely mimics the biological reality. Finally, we observed that if the value of *stableCycleDelay* was left unchanged the rate of increase of CELL number was too low. Thus, *stableCycleDelay* was lowered to allow for more CELL DIVISION and more CELL DEATH. It is possible that further empirical tuning of *cellCycle*, *deathRateLumen*, and *stableCycleDelay* could produce results that attain a

higher degree of similarity than reported above.

For the TS ISMA-C the value of *shiftDelay* that produced the greatest similarity was quite close to that for the TS ISMA-M (190 versus 200). However, because POLARIZATION is delayed in the ISMA-C, this value resulted in CELL stabilization at a later point in the TS ISMA-C versus the TS ISMA-M (6.4 DAYS versus 5.24 DAYS).

4.4.4. Parameters unchanged from the ISMA-M to the ISMA-C

We also reflected on the parameters that did not require changing to create the ISMA-C. These included *lumenGrowthRate* and *lgrSubtract*, *deathRateEpi,* and variables associated with mean CELL size. One of the most significant observations made within this report is that the variables that directly control the rate of LUMEN expansion (*lumenGrowthRate*) and the effect of CELL stretching on LUMEN expansion (*lgrSubtract*) are the same in the ISMA-M and ISMA-C, yet the two ISMAs mimicked the dramatically different lumen sizes observed in vitro. We believe this result is due to a combination of the later LUMEN initiation time and smaller number of CELLS within the ISMA-C, which allows for smaller LUMENS using the same operating principle and parameters.

We also noted during parameter tuning that it was not strictly necessary to alter the value of *deathRateEpi* to obtain an increased rate of CELL DEATH in the ISMA-C, likely due to the decreased value of *dyingShrinkRate*. Finally, we did not alter the variables that control mean CELL size, but the values within the ISMA-C were very similar to the values of mean CELL size observed in collagen. It does make sense that this would be the case, since cell size is similar over time in Matrigel and collagen, but mean cell size is somewhat different during the first three days of growth, a trend mimicked by the ISMA-M and ISMA-C.

4.4.5. Conlusion

The iterative process of modifying the ISMA to survive falsification by data from MDCK cystogenesis in collagen informed our understanding of the similarities and the differences in MDCK cystogenesis in Matrigel and collagen. While the qualitative aspects of cystogenesis are similar, including cell number increase, the formation of one or more lumens, and cyst expansion, the quantitative aspects are different. Cell polarization and lumen initiation occur later, cell number increases at a different rate, and cyst and lumen size are different. Cell size is similar within both systems, and stabilization occurs in both at around day 6, though slightly later in collagen.

The ISMA-C, which survives falsification against qualitative and quantitative data in collagen, provides additional evidence of the shared operating principles exhibited during MDCK cystogenesis in different culture systems. The two versions of the ISMA contain identical operating principles and computer code, differing only in a limited number of parameters. The ISMA-C demonstrates that delayed CELL POLARIZATION, which does contribute to the in silico phenotype, is only one aspect of the changes required to shift the ISMA to mimic MDCK cytogenesis in collagen. It also shows that it is not necessary to alter the rate of LUMEN expansion to achieve a reasonable measure of similarity, and that the differing lumen sizes observed in the different in vitro culture conditions are likely due to differences in lumen initiation time and cell number.

The ISMA-C provides a testable prediction of the behavior of MDCK cells in collagen when their axis of division is altered. According to the results of ISMA-C experiments, when the axis of cell division is randomized or reversed, overall quantitative cyst behavior will not change significantly, but only a tiny percentage of cysts will develop single lumens. The most important conclusion that can be derived from the

development and validation of the ISMA-C is that it further validates the ISMA and the specification used to achieve the targeted attributes. The ISMA-C survived falsification from a number of new quantitative measures of data, almost doubling the quantitative measures of similarity achieved. It provides significant evidence that MDCK cells in vitro follow the same set of operating principles in varied environmental conditions.

5. Summary

We presented an idea: under the conditions of epithelial cell growth in culture, molecular cell biology manifests at the cell level in what can described as a small set of operating principles that are responsible for the characteristic in vitro phenotypic attributes. We anticipated needing to identify and understand the operating principles in order to better understand how specific, detailed subcellular events may be linked to attributes of systemic EMT6 spheroid growth or MDCK cell cystogenesis. Our method and approach are diagrammed in Figure 1.1. We designed, refined, and tuned quasi-autonomous software components that, upon execution, formed abstract analogues. We showed that measures of behaviors during simulated growth were similar to available wet-lab data using quantitative similarity measures. We submit that our analogue mechanisms, with emphasis on the explicit AXIOMS, may stand as a plausible, abstract hypothesis for what was observed during in vitro growth experiments.

Because analogue components are quasi-autonomous, when the current set of targeted attributes is expanded one at a time, it is relatively straightforward to revise analogues to match each new, expanded set. We achieved the targeted attributes with SMS and ISMA CELLS that exist in three states (though the ISMA required LUMEN and MATRIX components as well). When the attribute lists are expanded (even to include pathological attributes of drug treatments), it is straightforward to add new cell states that possess different axiomatic operating principles. A future challenge will be to build a parallel system in which each (or some) atomic CELL component and its operating principles are replaced with a composite CELL object containing a set of interacting components intended to map to modular components within epithelial cells. During

INTRACELLULAR interaction, specific internal components would each use a portion of the same local environment information to act on other internal components such that actions are essentially identical to the current events that occur in vitro. The resulting growth characteristics would be indistinguishable from those described herein. The two systems could be iteratively advanced in parallel as new information and data were added to the set of targeted attributes.

While many avenues for further exploration using both the SMS and ISMAs present themselves, the most important may yet be undiscovered. This report demonstrated that agent-oriented models of disparate systems could be developed using a unified approach; build a list of targeted attributes describing a referent system, create a specification that might achieve those attributes, implement analogues that execute that specification, and then iteratively refine the analogues (potentially with new in vitro validating data). The resulting analogues and the process by which they are achieved can increase our knowledge of the underlying referent system and also of the process of modeling creation in and of itself. These analogues can be independent of modeling framework, programming language, in vitro referent, resolution and dimensionality, as is readily apparent when the SMS and ISMA are compared, which have almost none of these in common. Yet due to the similar approach taken, the lessons learned in the creation of the SMS could be successfully applied during the development of the ISMA.

Although the underlying approach evolved during the process, it was not necessary to begin, philosophically, from scratch. It is our hope that future modeling efforts, whether seeking to mimic behaviors of tumor spheroids, MDCK cystogenesis, or other novel phenomena, can utilize the approach detailed herein to significantly reduce

the time required for analogue creation and refinement. In addition, we anticipate that the clear contribution to the success of the ISMA from our active wet-lab collaboration will encourage readers to pursue similar collaborations. This report should provide ample evidence of the benefits to the wet-lab researcher of increased understanding, quantitative robustness, and efficient and economical hypothesis testing and generation. Whether the reader is focused on cell biological processes, morphogenesis, or other phenomena, we are confident that these techniques and results provide a strong philosophical foundation for the creation and revision of cell biological analogues.

6. References

1. Gove PB, editor. (2002) Webster's third new international dictionary, unabridged. Springfield, MA: Merriam-Webster.

2. Jiang Y, Pjesivac-Grbovic J, Cantrell C, Freyer J. (2005) A multiscale model for avascular tumor growth. Biophys J : 3884-3894.

3. Roose T, Chapman SJ, Maini PK. (2007) Mathematical models of avascular cancer. SIAM Review : 179-208.

4. Araujo RP, McElwain DLS. (2004) A history of the study of solid tumour growth: The contribution of mathematical modelling. Bull Math Biol : 1039-1091.

5. Grant MR, Mostov KE, Tlsty TD, Hunt CA. (2006) Simulating properties of in vitro epithelial cell morphogenesis. PLoS Comput Biol 2: e129.

6. Schaller G, Meyer-Hermann M. (2005) Multicellular tumor spheroid in an off-lattice voronoi-delaunay cell model. Phys Rev E Stat Nonlin Soft Matter Phys : 051910-051935.

7. Tang J, Hunt CA. (2010) Identifying the rules of engagement enabling leukocyte rolling, activation, and adhesion. PLoS Comput Biol 6: e1000681.

8. Burton AC. (1966) Rate of growth of solid tumours as a problem of diffusion. Growth : 157-176.

9. Casciari JJ, Sotirchos SV, Sutherland RM. (1992) Mathematical modelling of microenvironment and growth in EMT6/Ro multicellular tumour spheroids. Cell Prolif : 1-22.

10. Chignola R, Milotti E. (2005) A phenomenological approach to the simulation of metabolism and proliferation dynamics of large tumour cell populations. Phys Biol : 8-22.

11. Kim Y, Stolarska M, Othmer HG. (2007) A hybrid model for tumor spheroid growth in vitro I: Theoretical development and early results. Math Models Methods Appl Sci : 1773-1798.

12. Galle J, Loeffler M, Drasdo D. (2005) Modeling the effect of deregulated proliferation and apoptosis on the growth dynamics of epithelial cell populations in vitro. Biophys J : 62-75.

13. Anderson ARA, Weaver AM, Cummings PT, Quaranta V. (2006) Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell : 905-915.

14. Rejniak KA, Anderson AR. (2008) A computational study of the development of epithelial acini: I. sufficient conditions for the formation of a hollow structure. Bull Math Biol 70: 677-712.

15. Kim SH, Park S, Mostov K, Debnath J, Hunt CA. (2009) Computational investigation of epithelial cell dynamic phenotype in vitro. Theor Biol Med Model 6: 8.

16. Fisher J, Henzinger TA. (2007) Executable cell biology. Nat Biotech : 1239-1249.

17. Hunt CA, Ropella GEP, Park S, Engelberg JA. (2008) Dichotomies between computational and mathematical models. Nat Biotech : 737-738.

18. Longo D, Peirce SM, Skalak TC, Davidson L, Marsden M, et al. (2004) Multicellular computer simulation of morphogenesis: Blastocoel roof thinning and matrix assembly in xenopus laevis. Dev Biol : 210-222.

19. Christley S, Alber MS, Newman SA. (2007) Patterns of mesenchymal condensation in a multiscale, discrete stochastic model. PLoS Comput Biol 3: e76.

20. Yan L, Ropella GEP, Park S, Roberts MS, Hunt CA. (2008) Modeling and simulation of hepatic drug disposition using a physiologically based, multi-agent in silico liver. Pharm Res : 1023-1036.

21. Freyer JP, Sutherland RM. (1986) Regulation of growth saturation and development of necrosis in EMT6/Ro multicellular spheroids by the glucose and oxygen supply. Cancer Res : 3504-3512.

22. Freyer JP. (1988) Role of necrosis in regulating the growth saturation of multicellular spheroids. Cancer Res : 2432-2439.

23. Sutherland RM, Durand RE. (1976) Radiation response of multicell spheroidsÑan in vitro tumour model. Curr Top Radiat Res Q : 87-139.

24. Freyer JP, Sutherland RM. (1980) Selective dissociation and characterization of cells from different regions of multicell tumor spheroids. Cancer Res : 3956-3965.

25. Freyer JP, Sutherland RM. (1985) A reduction in the in situ rates of oxygen and glucose consumption of cells in EMT6/Ro spheroids during growth. J Cell Physiol : 516-524.

26. Freyer JP, Sutherland RM. (1986) Proliferative and clonogenic heterogeneity of cells from EMT6/Ro multicellular spheroids induced by the glucose and oxygen supply. Cancer Res : 3513-3520.

27. Leaf C. (2004) Why we're losing the war on cancer (and how to win it). Fortune : 76-8.

28. LaRue KEA, Khalil M, Freyer JP. (2004) Microenvironmental regulation of proliferation in multicellular spheroids is mediated through differential expression of cyclin-dependent kinase inhibitors. Cancer Res : 1621-1631.

29. Wartenberg M, Acker H. (1995) Quantitative recording of vitality patterns in living multicellular spheroids by confocal microscopy. Micron : 395-404.

30. Chignola R, Schenetti A, Andrighetto G, Chiesa E, Foroni R, et al. (2000) Forecasting the growth of multicell tumour spheroids: Implications for the dynamic growth of solid tumours. Cell Prolif : 219-229.

31. Luke S, Balan GC, Panait LA, Paus S. (2003) MASON: A java multiagent simulation library. Proceedings of Agent 2003 Conference on Challenges in Social Simulation : 49-64.

32. Freyer JP. (1988) Role of necrosis in regulating the growth saturation of multicellular spheroids. Cancer Res 48: 2432-2439.

33. Harris LK, Keogh RJ, Wareing M, Baker PN, Cartwright JE, et al. (2006) Invasive trophoblasts stimulate vascular smooth muscle cell apoptosis by a fas ligand-dependent mechanism. Am J Pathol : 1863-1874.

34. Dorie MJ, Kallman RF, Rapachhietta DF, Van-Antwerp D, Huang YR. (1982) Migration and internalization of cells and polystyrene microspheres in tumor cell spheroids. Exp Cell Res : 201-209.

35. Landry J, Freyer JP, Sutherland RM. (1981) Shedding of mitotic cells from the surface of multicell spheroids during growth. J Cell Physiol : 23-32.

36. Casciari JJ, Sotirchos SV, Sutherland RM. (1992) Variations in tumor cell growth rates and metabolism with oxygen concentration, glucose concentration, and extracellular pH. J Cell Physiol : 386-394.

37. Walker DC, Southgate J, Hill G, Holcombe M, Hose DR, et al. (2004) The epitheliome% agent-based modelling of the social behaviour of cells. BioSystems : 89-100.

38. Ermentrout GB, Edelstein-Keshet L. (1993) Cellular automata approaches to biological modeling. J Theor Biol 160: 97-133.

39. Dormann S, Deutsch A. (2002) Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. In Silico Biol : 393-406.

40. Rejniak KA. (2005) A single-cell approach in modeling the dynamics of tumor microregions. Math Biosc Eng : 643-655.

41. Witten TA, Sander LM. (1981) Diffusion-limited aggregation, a kinetic critical phenomenon. Phys Rev Lett : 1400-1403.

42. Engelberg JA, Hunt CA. (2007) A local mechanism generates saturation in an in silico model of in vitro multicellular tumor spheroid growth. J Crit Care : 331-331.

43. Gompertz B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society of London : 513-583.

44. Delsanto PP, Condat CA, Pugno N, Gliozzi AS, Griffa M. (2008) A multilevel approach to cancer growth modeling. J Theor Biol : 16-24.

45. Kim SH, Debnath J, Mostov K, Park S, Hunt CA. (2009) A computational approach to resolve cell level contributions to early glandular epithelial cancer progression. BMC Syst Biol 3: 122.

46. Wang AZ, Ojakian GK, Nelson WJ. (1990) Steps in the morphogenesis of a polarized epithelium. I. uncoupling the roles of cell-cell and cell-substratum contact in establishing plasma membrane polarity in multicellular epithelial (MDCK) cysts. J Cell Sci 95: 137-151.

47. Zheng Z, Zhu H, Wan Q, Liu J, Xiao Z, et al. (2010) LGN regulates mitotic spindle orientation during epithelial morphogenesis. J Cell Biol 189: 275-288.

48. O'Brien LE, Jou TS, Pollack AL, Zhang Q, Hansen SH, et al. (2001) Rac1 orientates epithelial apical polarity through effects on basolateral laminin assembly. Nat Cell Biol 3: 831-838.

49. Yu W, Datta A, Leroy P, O'Brien LE, Mak G, et al. (2005) Beta1-integrin orients epithelial polarity via Rac1 and laminin. Mol Biol Cell 16: 433-445.

50. Kim M, Datta A, Brakeman P, Yu W, Mostov KE. (2007) Polarity proteins PAR6 and aPKC regulate cell death through GSK-3beta in 3D epithelial morphogenesis. J Cell Sci 120: 2309-2317.

51. Martín-Belmonte F, Yu W, Rodriguez-Fraticelli AE, Ewald AJ, Werb Z, et al. (2008) Cell-polarity dynamics controls the mechanism of lumen formation in epithelial morphogenesis. Curr Biol 18: 507-513.

52. Cickovski TM, Huang C, Chaturvedi R, Glimm T, Hentschel HG, et al. (2005) A framework for three-dimensional simulation of morphogenesis. IEEE/ACM Trans Comput Biol Bioinform 2: 273-288.

53. Rejniak KA, Wang SE, Bryce NS, Chang H, Parvin B, et al. (2010) Linking changes in epithelial morphogenesis to cancer mutations using computational modeling. PLoS Comput Biol 6: e1000900.

54. Engelberg JA, Ropella GE, Hunt CA. (2008) Essential operating principles for tumor spheroid growth. BMC Syst Biol 2: 110.

55. Kim SH, Matthay MA, Mostov K, Hunt CA. (2010) Simulation of lung alveolar epithelial wound healing in vitro. J R Soc Interface 7: 1157-1170.

56. Lam TN, Hunt CA. (2010) Mechanistic insight from in silico pharmacokinetic experiments: Roles of P-glycoprotein, Cyp3A4 enzymes, and microenvironments. J Pharmacol Exp Ther 332: 398-412.

57. Hunt CA, Ropella GE, Lam TN, Tang J, Kim SH, et al. (2009) At the biological modeling and simulation frontier. Pharm Res 26: 2369-2400.

58. Kim SH, Yu W, Mostov K, Matthay MA, Hunt CA. (2009) A computational approach to understand in vitro alveolar morphogenesis. PLoS One 4: e4819.

59. Swat MH, Hester SD, Heiland RW, Zaitlen BL, Glazier JA, et al. (2010) CompuCell3D manual, version 3.4.2. 2010: 123.

60. Glazier JA, Balter A, Popławski NJ. (2007) Magnetization to morphogenesis: A brief history of the glazier-graner-hogeweg model. In: Rejniak KA, editor. Single-Cell-Based Models in Biology and Medicine. : Birkhäuser Basel. pp. 79-106.

61. Graner F, Glazier JA. (1992) Simulation of biological cell sorting using a two-dimensional extended potts model. Phys Rev Lett 69: 2013-2016.

62. Nathke IS, Hinck L, Swedlow JR, Papkoff J, Nelson WJ. (1994) Defining interactions and distributions of cadherin and catenin complexes in polarized epithelial cells. J Cell Biol 125: 1341-1352.

63. Collins JA, Schandi CA, Young KK, Vesely J, Willingham MC. (1997) Major DNA fragmentation is a late event in apoptosis. J Histochem Cytochem 45: 923-934.

64. Gavrieli Y, Sherman Y, Ben-Sasson SA. (1992) Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. J Cell Biol 119: 493-501.

65. Tyas L, Brophy VA, Pope A, Rivett AJ, Tavare JM. (2000) Rapid caspase-3 activation during apoptosis revealed using fluorescence-resonance energy transfer. EMBO Rep 1: 266-270.

66. O'Brien LE, Zegers MM, Mostov KE. (2002) Opinion: Building epithelial architecture: Insights from three-dimensional culture models. Nat Rev Mol Cell Biol 3: 531-537.

67. Martín-Belmonte F, Gassama A, Datta A, Yu W, Rescher U, et al. (2007) PTEN-mediated apical segregation of phosphoinositides controls epithelial morphogenesis through Cdc42. Cell 128: 383-397.

68. Lam TN, Hunt CA. (2009) Discovering plausible mechanistic details of hepatic drug interactions. Drug Metab Dispos 37: 237-246.

69. Campisi J, d'Adda di Fagagna F. (2007) Cellular senescence: When bad things happen to good cells. Nat Rev Mol Cell Biol 8: 729-740.

70. Wang Y, Liang Y, Vanhoutte PM. (2010) SIRT1 and AMPK in regulating mammalian senescence: A critical review and a working model. FEBS Lett .

71. Narala SR, Allsopp RC, Wells TB, Zhang G, Prasad P, et al. (2008) SIRT1 acts as a nutrient-sensitive growth suppressor and its loss is associated with increased AMPK and telomerase activity. Mol Biol Cell 19: 1210-1219.

72. Zu Y, Liu L, Lee MY, Xu C, Liang Y, et al. (2010) SIRT1 promotes proliferation and prevents senescence through targeting LKB1 in primary porcine aortic endothelial cells. Circ Res 106: 1384-1393.

73. Etienne-Manneville S. (2004) Cdc42--the centre of polarity. J Cell Sci 117: 1291-1300.

74. Martín-Belmonte F, Gassama A, Datta A, Yu W, Rescher U, et al. (2007) PTEN-mediated apical segregation of phosphoinositides controls epithelial morphogenesis through Cdc42. Cell 128: 383-397.

75. Bryant DM, Datta A, Rodriguez-Fraticelli AE, Peranen J, Martín-Belmonte F, et al. (2010) A molecular network for de novo generation of the apical surface and lumen. Nat Cell Biol 12: 1035-1045.

76. Rohatgi R, Ma L, Miki H, Lopez M, Kirchhausen T, et al. (1999) The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. Cell 97: 221-231.

77. Zhang H, Macara IG. (2008) The PAR-6 polarity protein regulates dendritic spine morphogenesis through p190 RhoGAP and the rho GTPase. Dev Cell 14: 216-226.

78. Warner SJ, Longmore GD. (2009) Cdc42 antagonizes Rho1 activity at adherens junctions to limit epithelial cell apical tension. J Cell Biol 187: 119-133.

79. Lin HH, Yang TP, Jiang ST, Yang HY, Tang MJ. (1999) Bcl-2 overexpression prevents apoptosis-induced madin-darby canine kidney simple epithelial cyst formation. Kidney Int 55: 168-178.

80. Yu W, Fang X, Ewald A, Wong K, Hunt CA, et al. (2007) Formation of cysts by alveolar type II cells in three-dimensional culture reveals a novel mechanism for epithelial morphogenesis. Mol Biol Cell 18: 1693-1700.

81. Ferrari A, Veligodskiy A, Berge U, Lucas MS, Kroschewski R. (2008) ROCK-mediated contractility, tight junctions and channels contribute to the conversion of a preapical patch into apical surface during isochoric lumen initiation. J Cell Sci 121: 3649-3663.

82. Ewald AJ, Brenot A, Duong M, Chan BS, Werb Z. (2008) Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. Dev Cell 14: 570-581.

83. Paszek MJ, Zahir N, Johnson KR, Lakins JN, Rozenberg GI, et al. (2005) Tensional homeostasis and the malignant phenotype. Cancer Cell 8: 241-254.

84. Wells RG. (2008) The role of matrix stiffness in regulating cell behavior. Hepatology 47: 1394-1400.

85. Marmaras A, Berge U, Ferrari A, Kurtcuoglu V, Poulikakos D, et al. (2010) A mathematical method for the 3D analysis of rotating deformable systems applied on lumen-forming MDCK cell aggregates. Cytoskeleton 67: 224-240.

86. Pollack AL, Runyan RB, Mostov KE. (1998) Morphogenetic mechanisms of epithelial tubulogenesis: MDCK cell polarity is transiently rearranged without loss of cell-cell contact during scatter factor/hepatocyte growth factor-induced tubulogenesis. Dev Biol 204: 64-79.

87. O'Brien LE, Yu W, Tang K, Jou TS, Zegers MM, et al. (2006) Morphological and biochemical analysis of Rac1 in three-dimensional epithelial cell cultures. Methods Enzymol 406: 676-691.

7. Appendix 1.

Parameter sweeping results.
Experiments followed the same design as in Figures 2 and 3. Designations are the same as in Figures 2 and 3. Parameters changed from settings in Table 2 are listed within each Figure. Figure 1 to 5 used the TS ISMA. Figure 6 through 80 used the LS ISMA. Except for S11-1 to S11-5, all parameters were fixed except the single parameter being varied.

Figure 1 to 5: TS ISMA.

Figure 6 to 9: varied *wedgeArea.*

Figure 10 to 13: varied *lambdaArea*.

Figure 14 to 17: varied *stableTargetArea*.

Figure 18 to 23: varied *cellCycle*.

Figure 24 to 27: varied *stableCycleDelay*.

Figure 28 to 31: varied *lambdaPerim*.

Figure 32 to 35: varied *polarDelay*.

Figure 36 to 39: varied *shiftDelay* with high *stableRatio*.

Figure 40 to 44: varied *lgrSubtract*.

Figure 45 to 48: varied *doublingArea.*

Figure 49 to 52: varied *multiplier*.

Figure 53 to 60: varied *lumenGrowthRate.*

Figure 61 to 64: varied *deathRateLumen*.

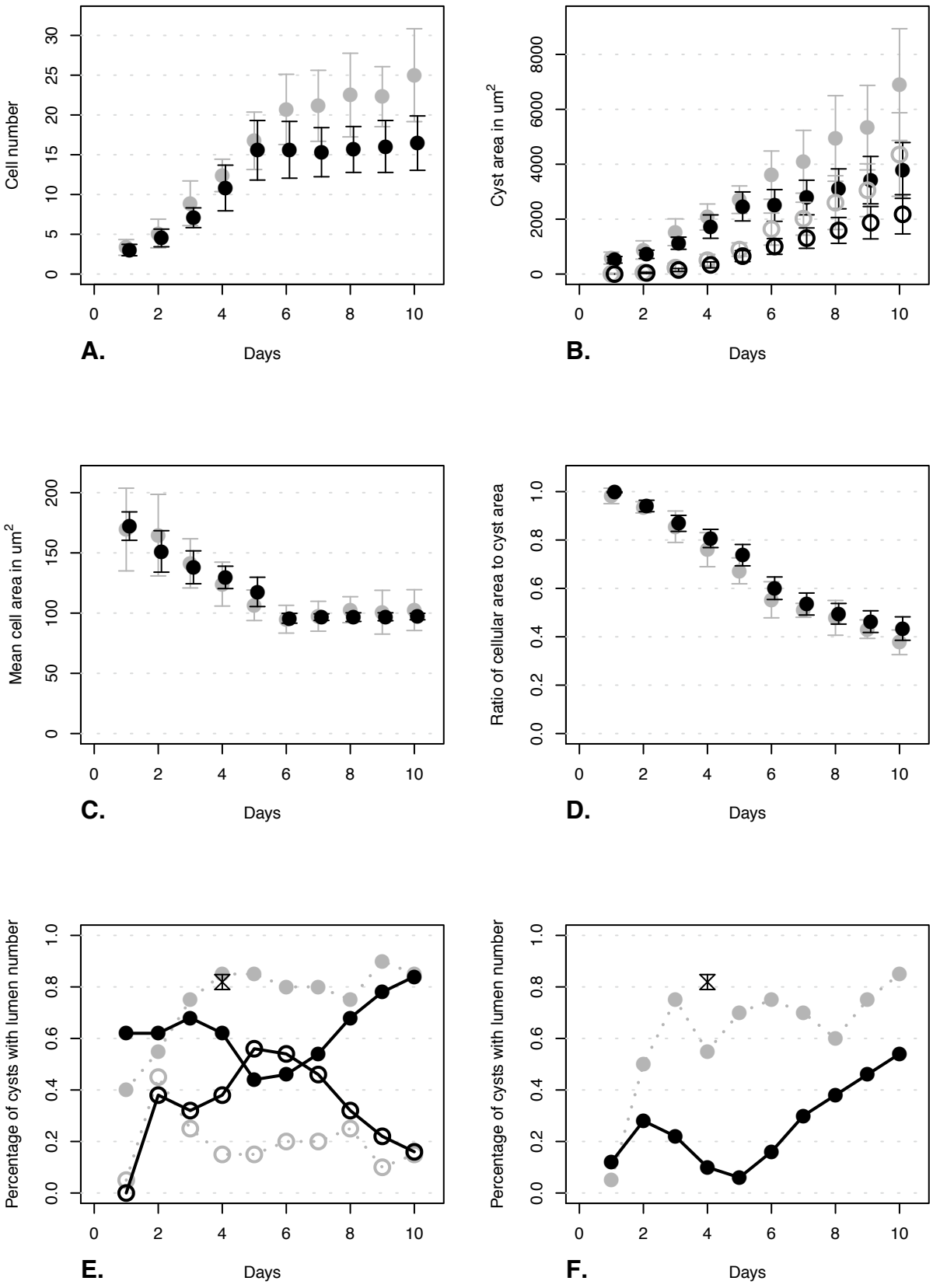
Figure 65 to 68: varied *deathRateEpi*.

Figure 69 to 72: varied *dyingShrinkRate*.

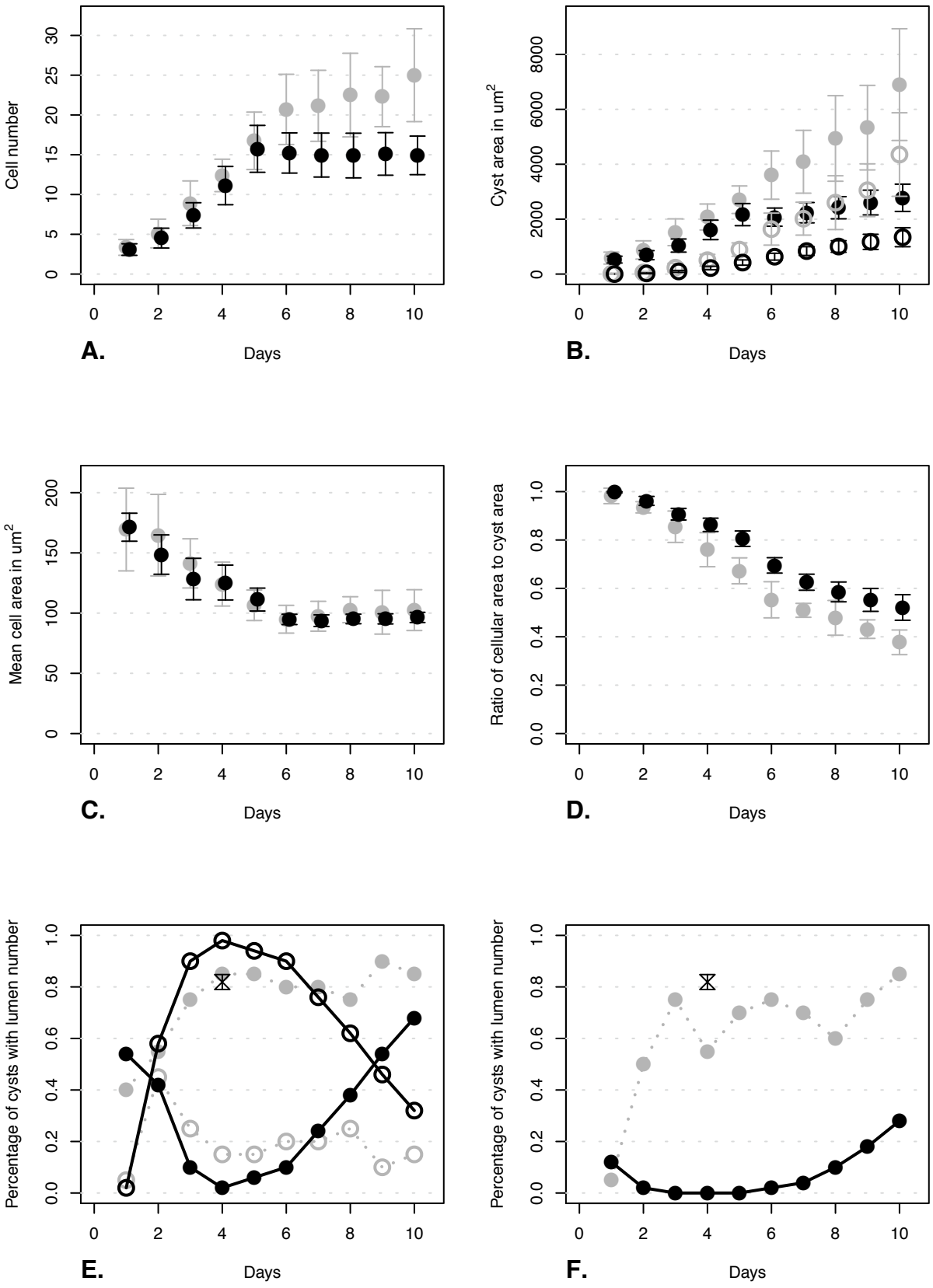
Figure 73 to 76: varied *clusterProb*.

Figure 77 to 80: varied *stableRatio*.

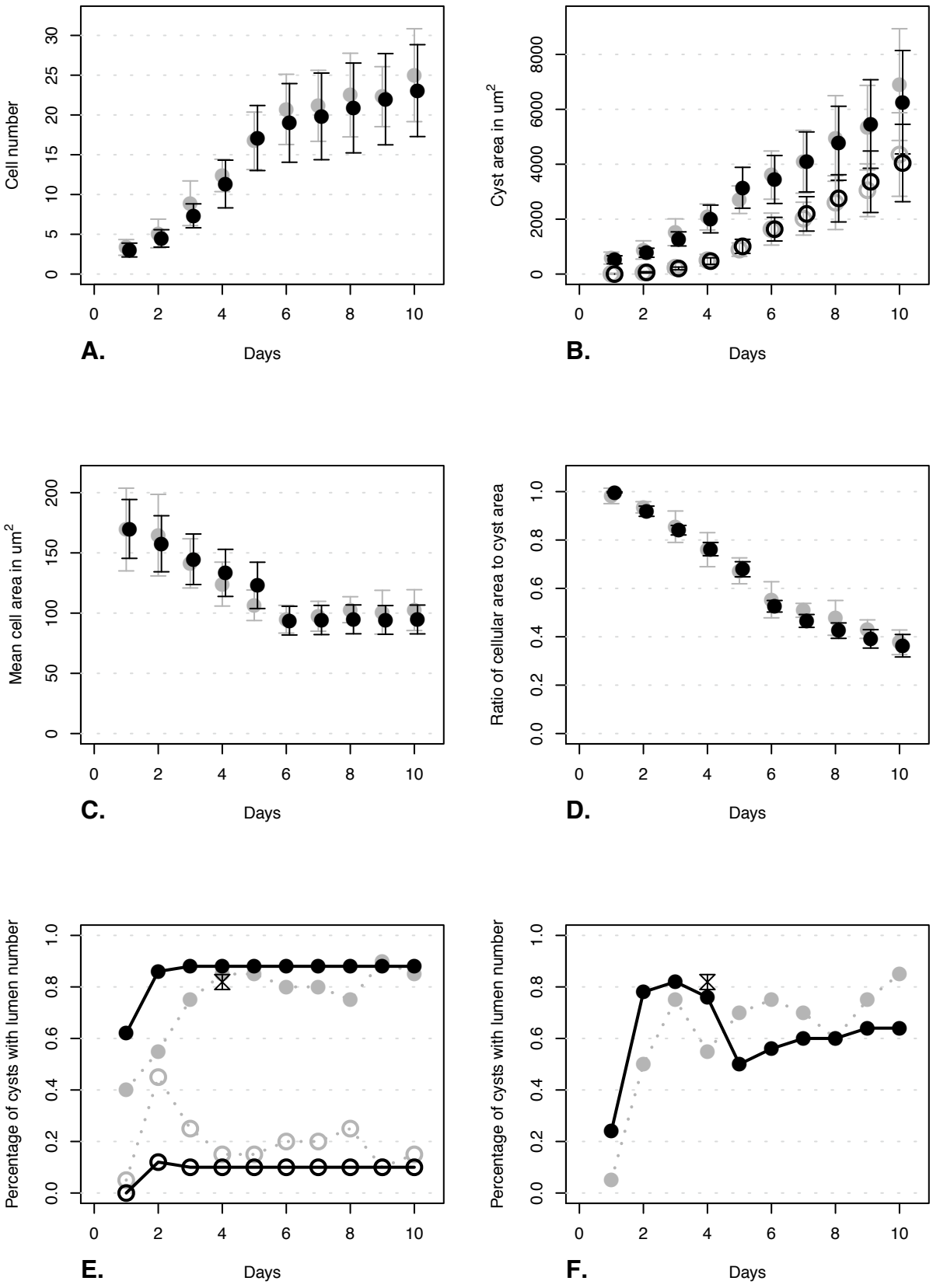
Appendix 1, Figure 1. TS ISMA with random DIVISION



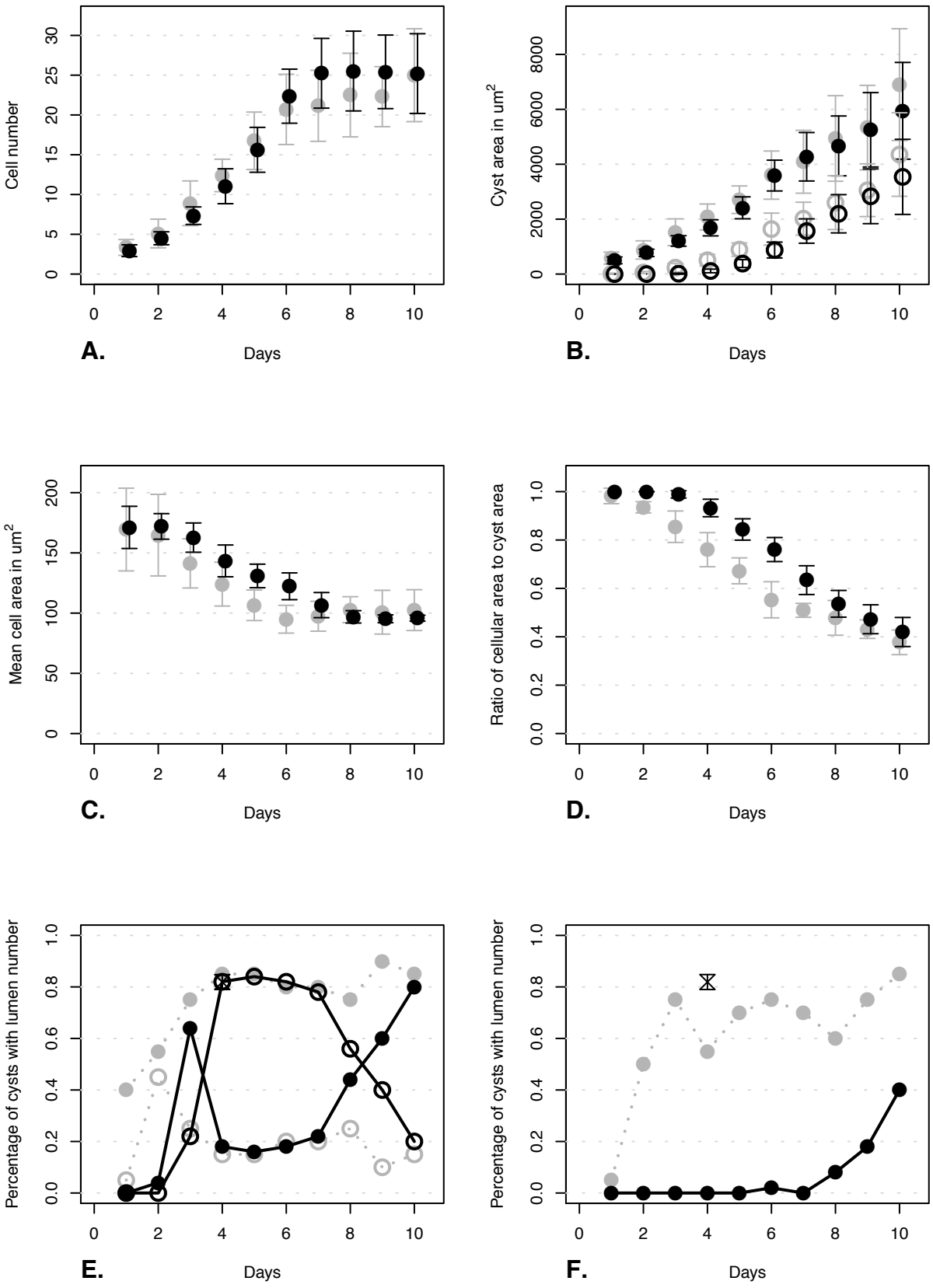
Appendix 1, Figure 2. TS ISMA with reversed DIVISION



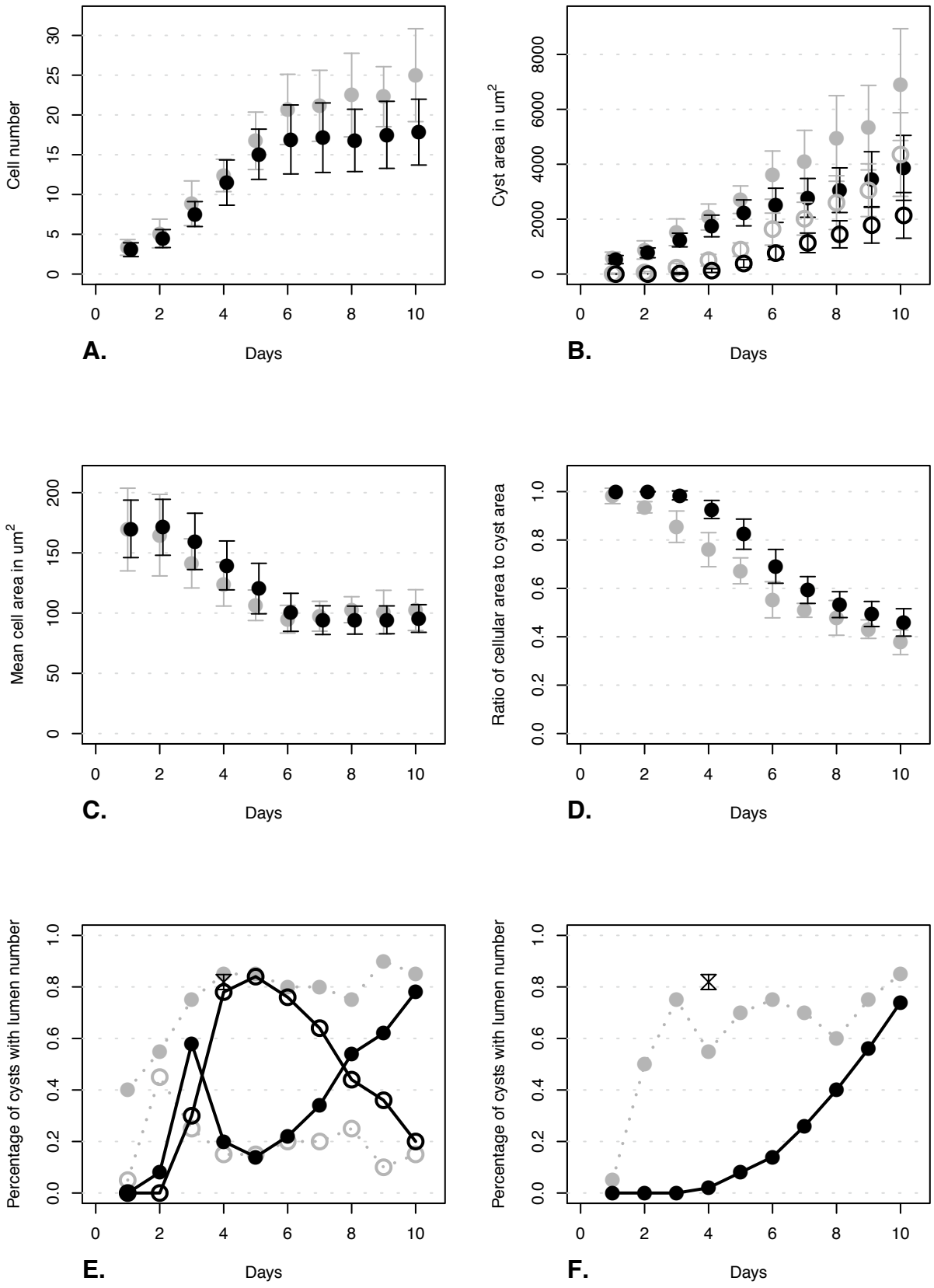
Appendix 1, Figure 3. TS ISMA with no LUMINAL CELL DEATH



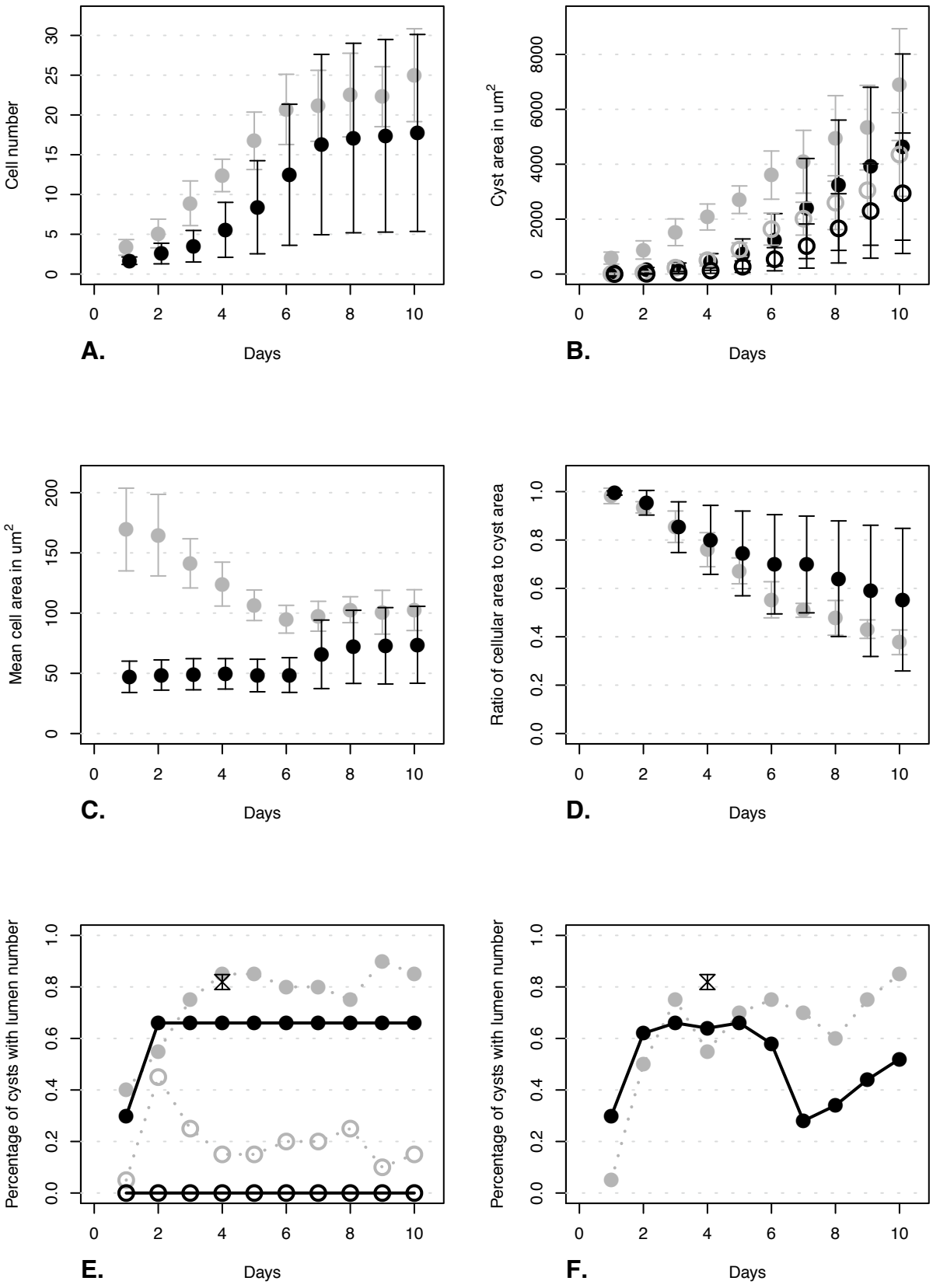
Appendix 1, Figure 4. TS ISMA with polarDelay = 130



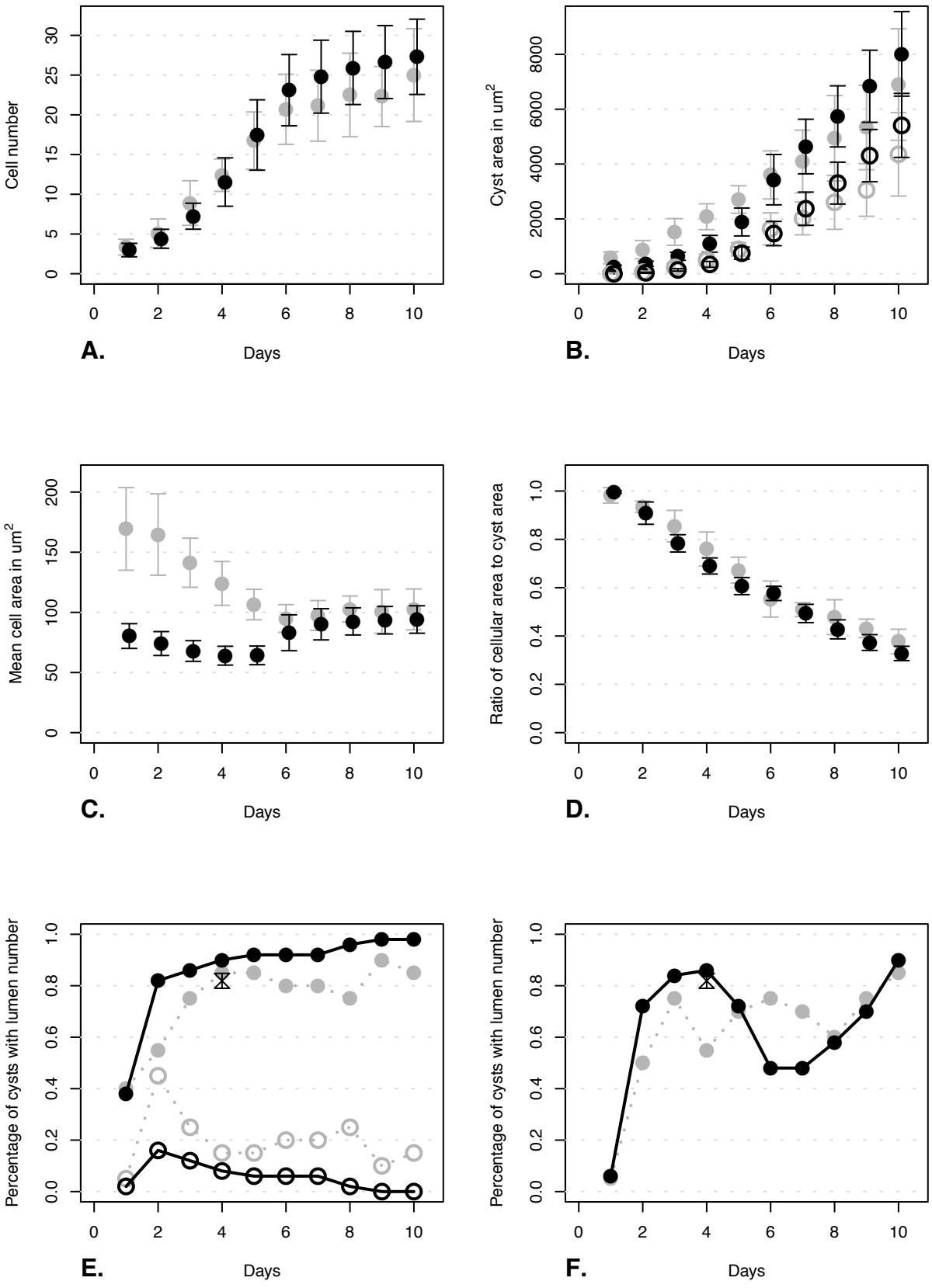
Appendix 1, Figure 5. TS ISMA with polarDelay = 130 and shiftDelay = 140



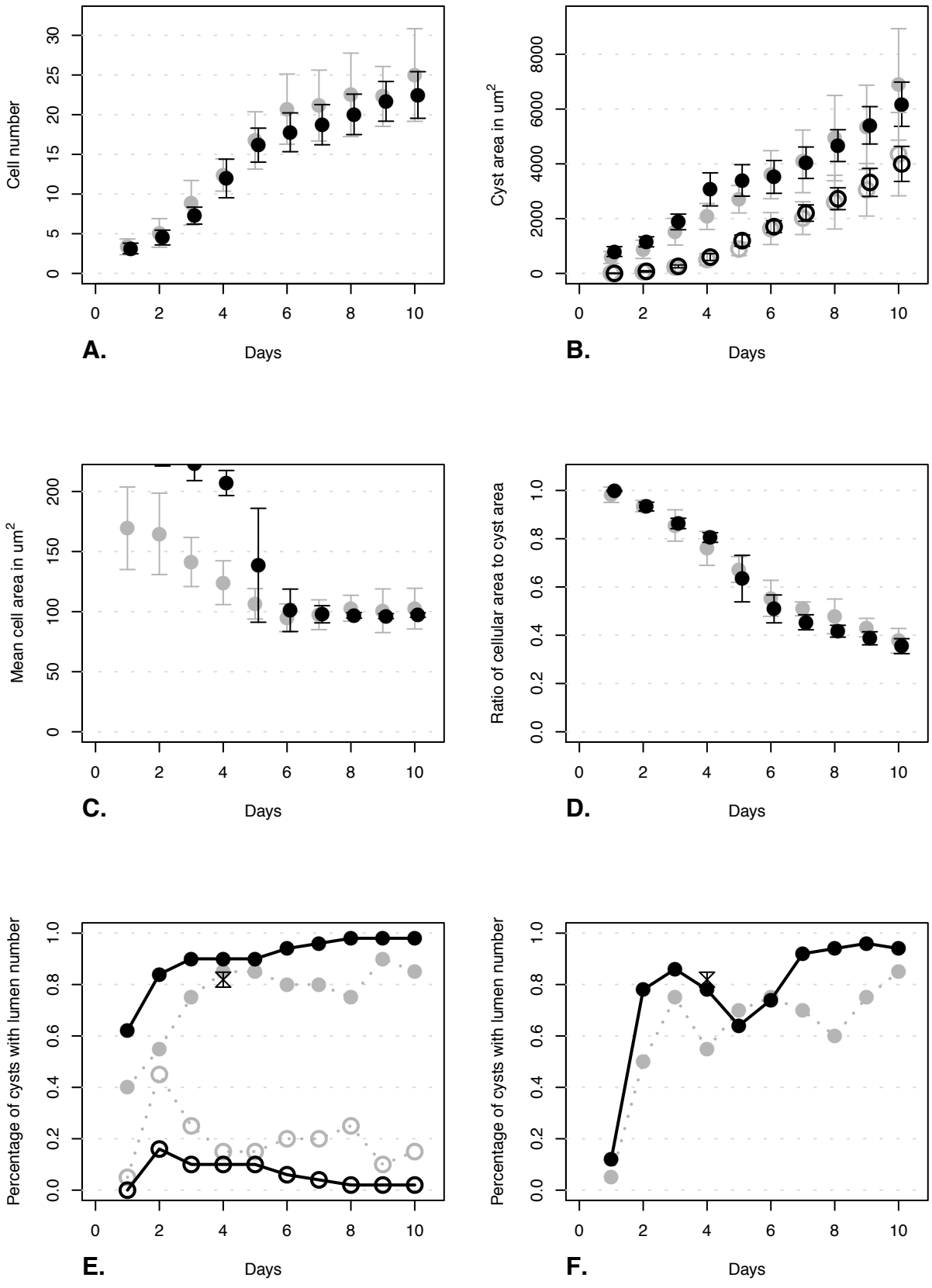
Appendix 1, Figure 6. Varied wedgeArea: w = 20



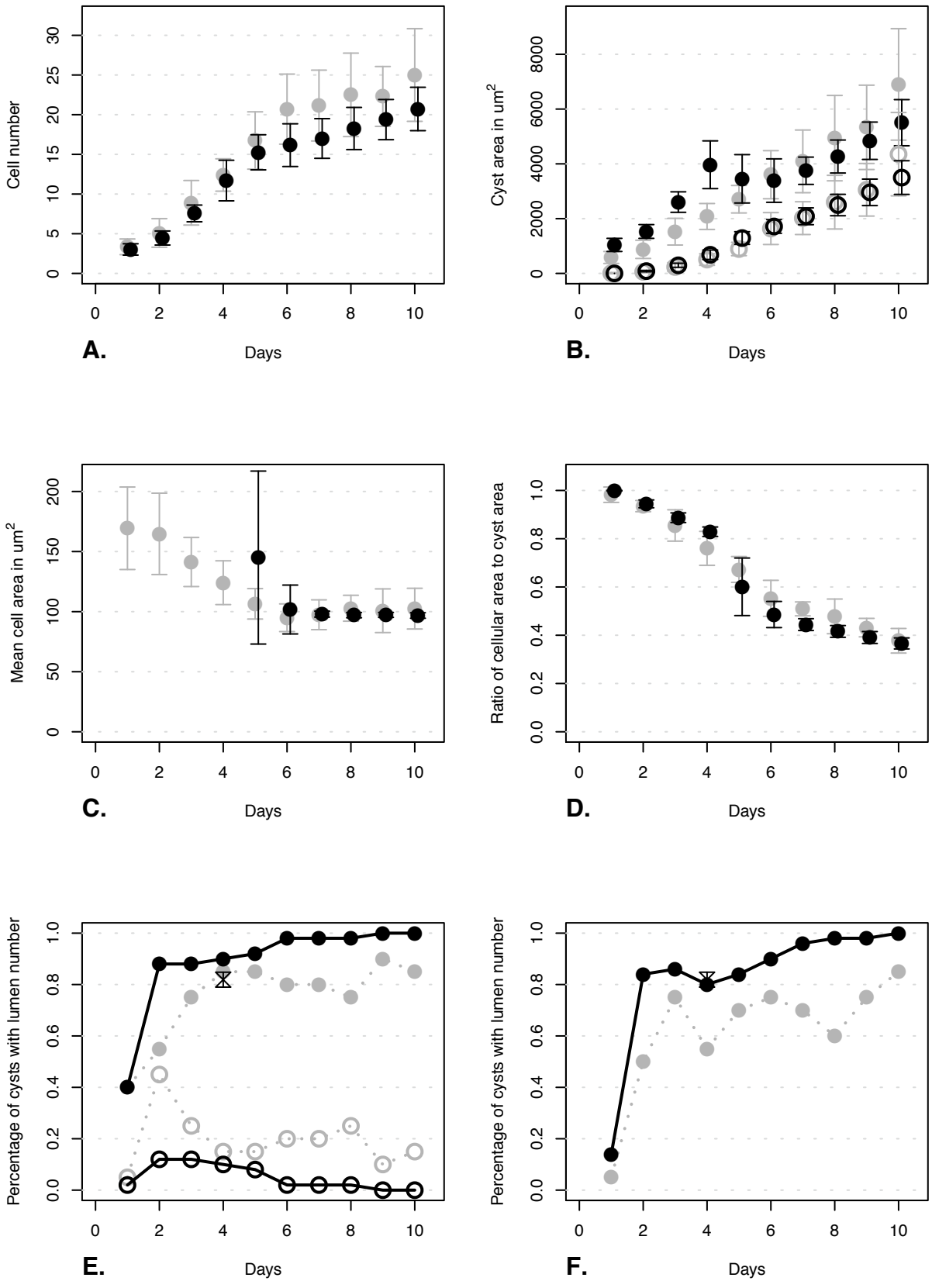
Appendix 1, Figure 7. Varied wedgeArea: w = 40



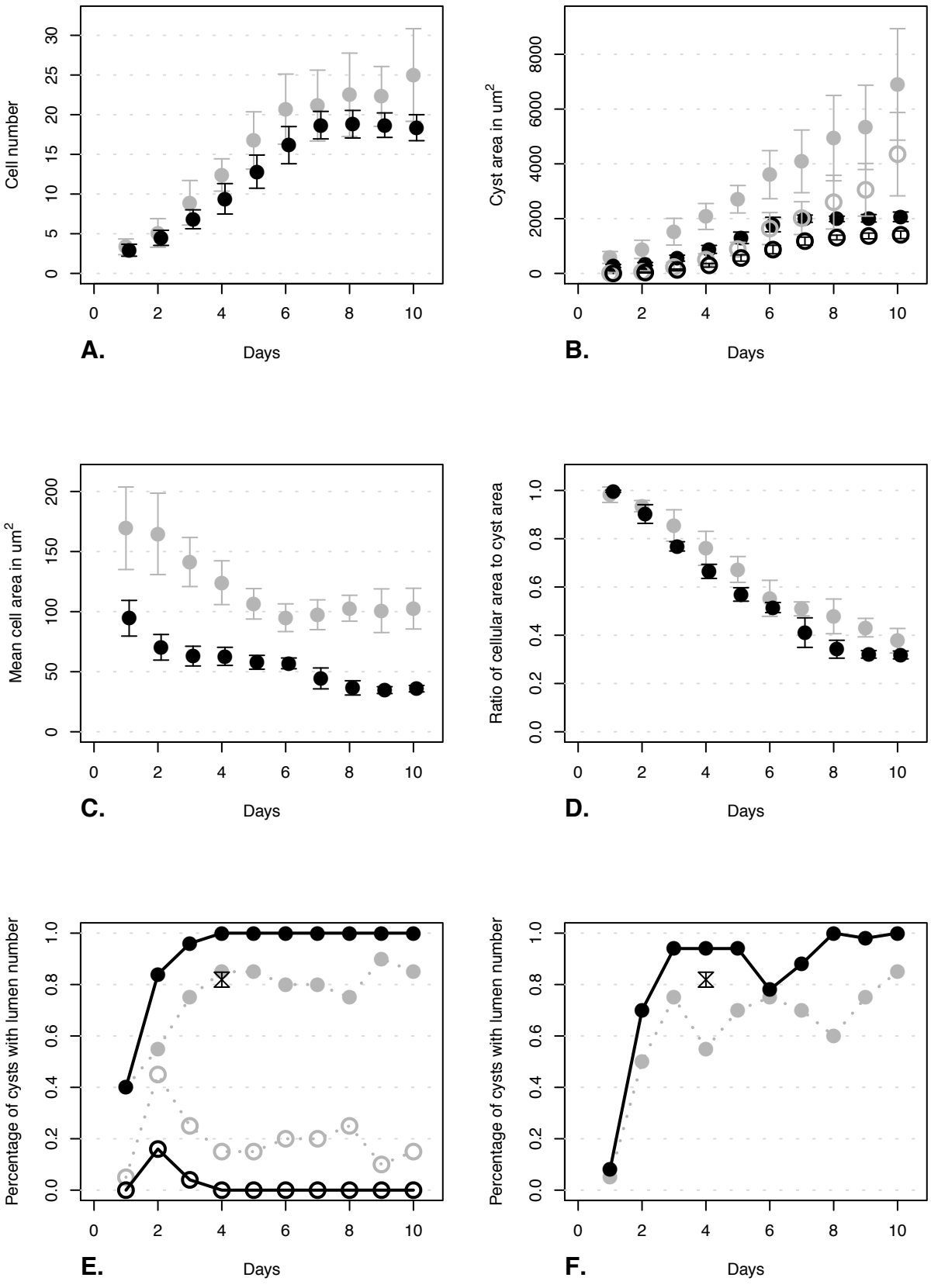
Appendix 1, Figure 8. Varied wedgeArea: w = 120



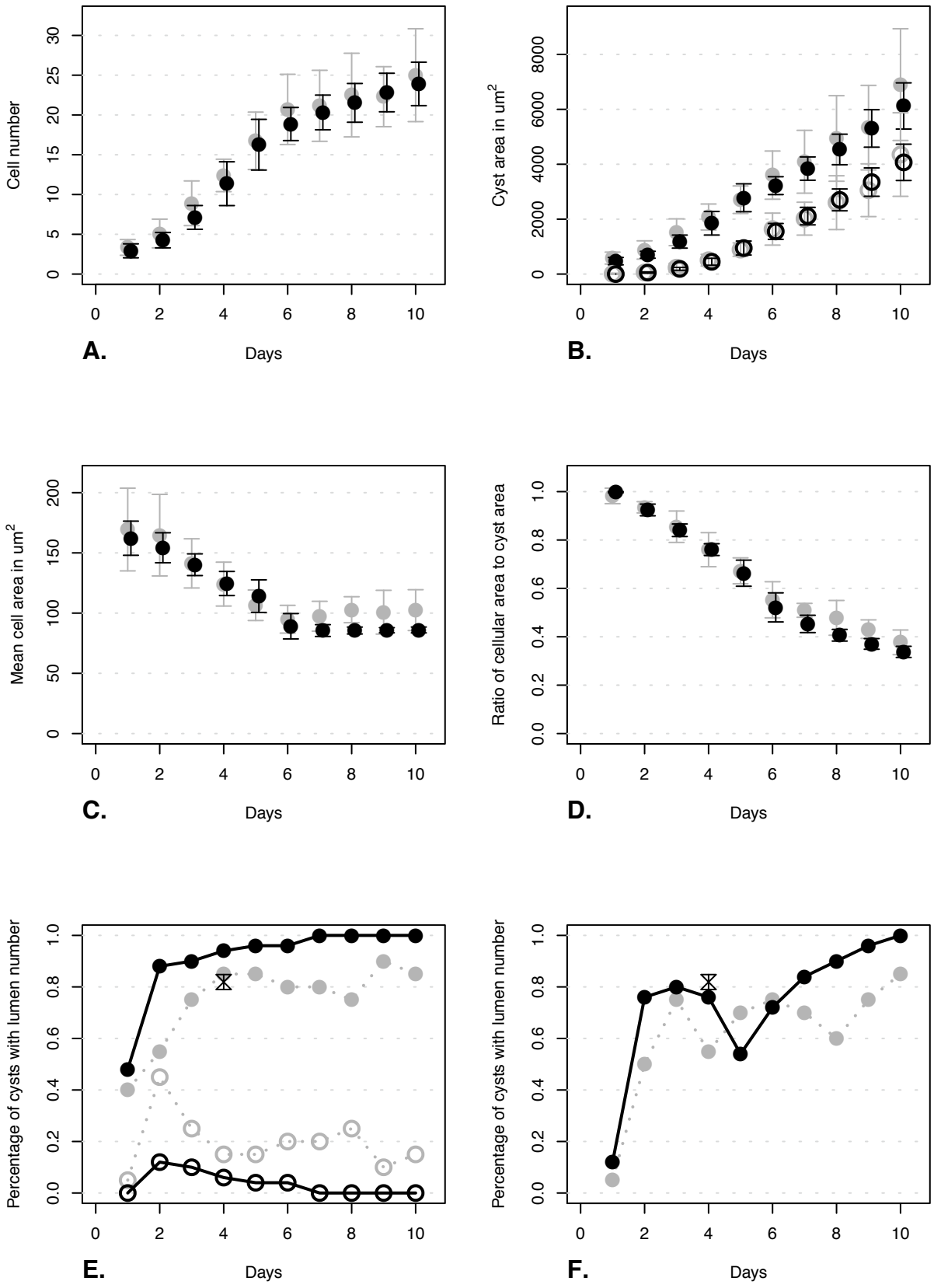
Appendix 1, Figure 9. Varied wedgeArea: w = 160



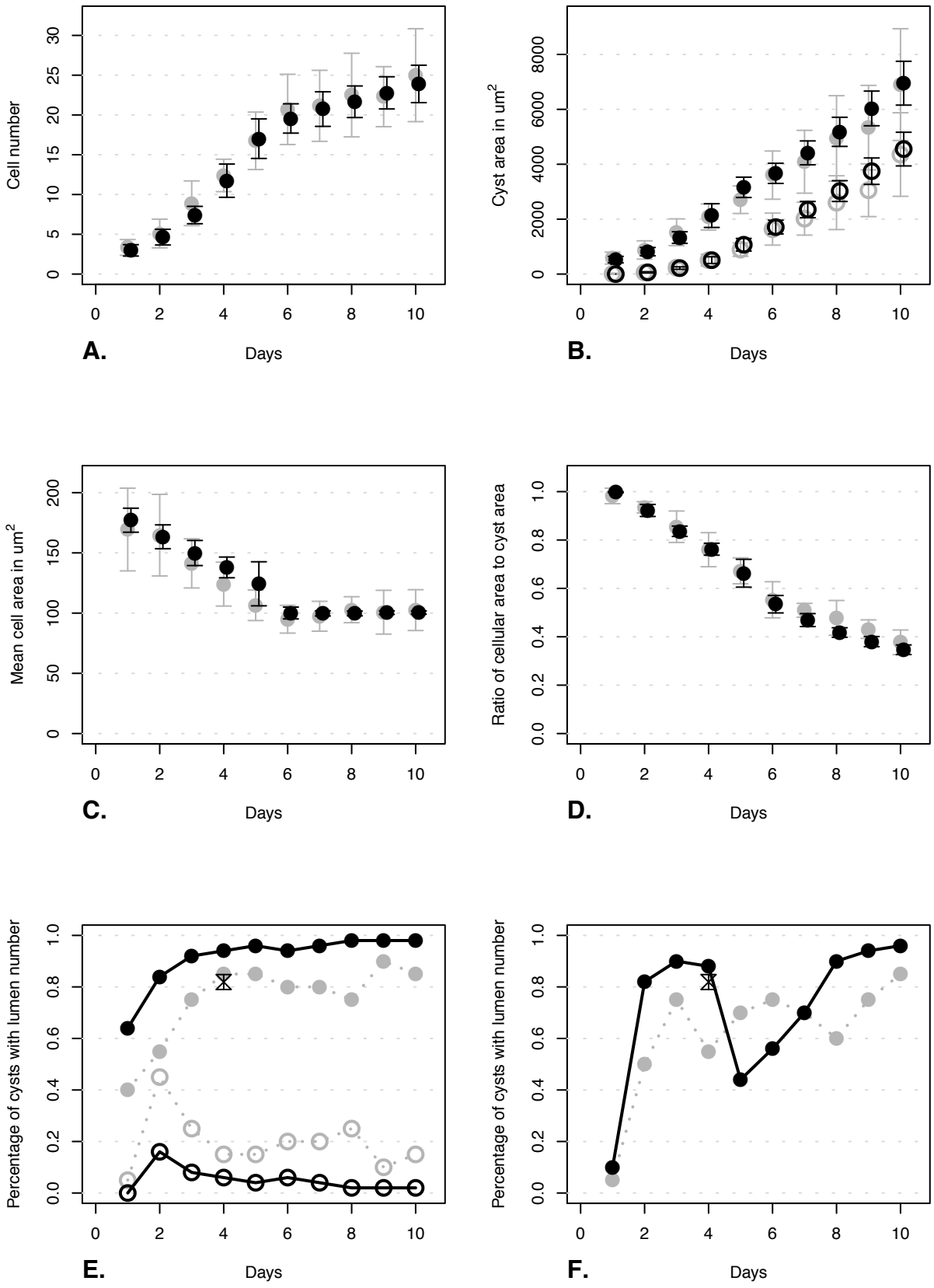
Appendix 1, Figure 10. Varied lambdaArea: la = 0.5



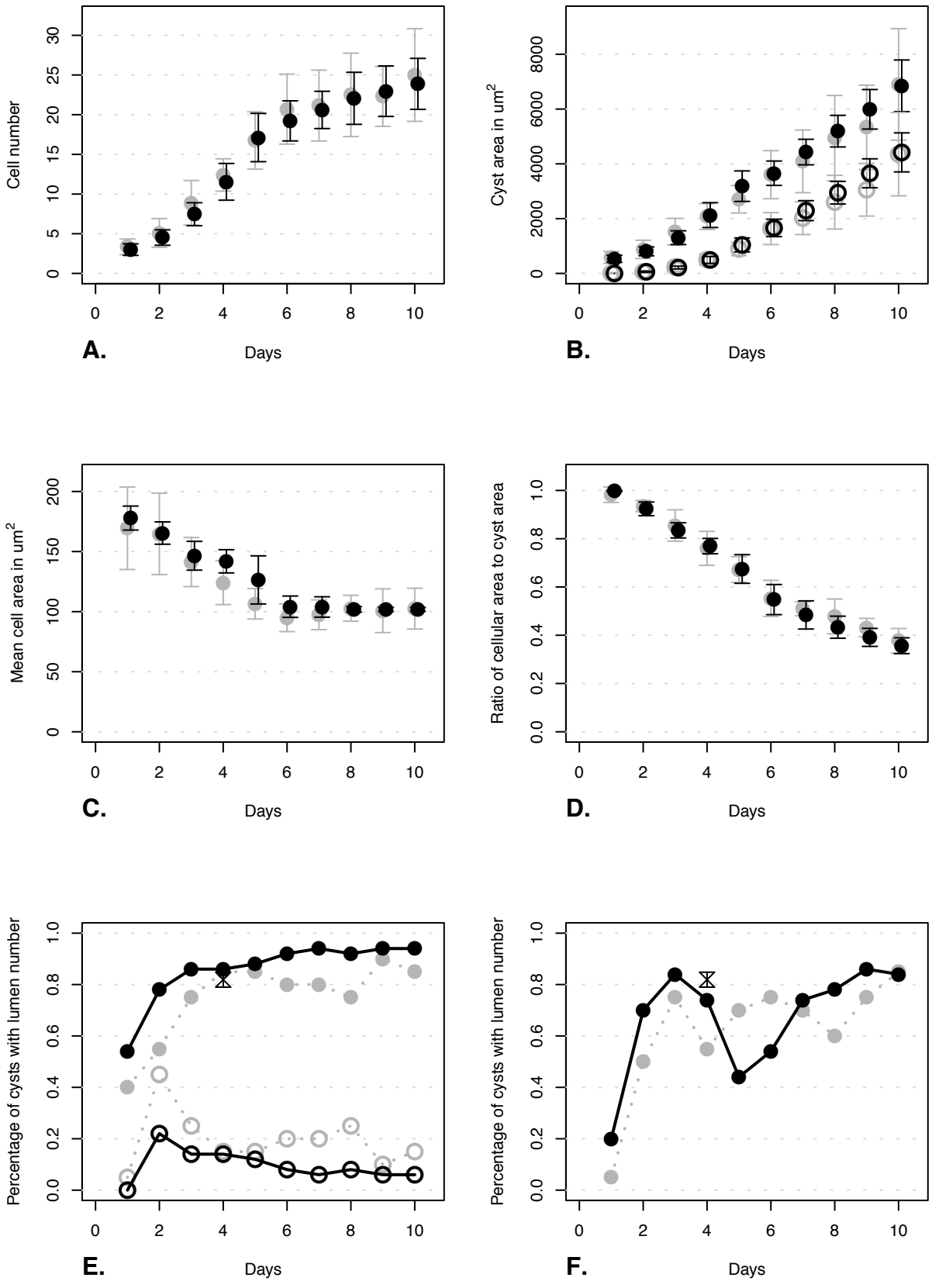
Appendix 1, Figure 11. Varied lambdaArea: la = 2.5



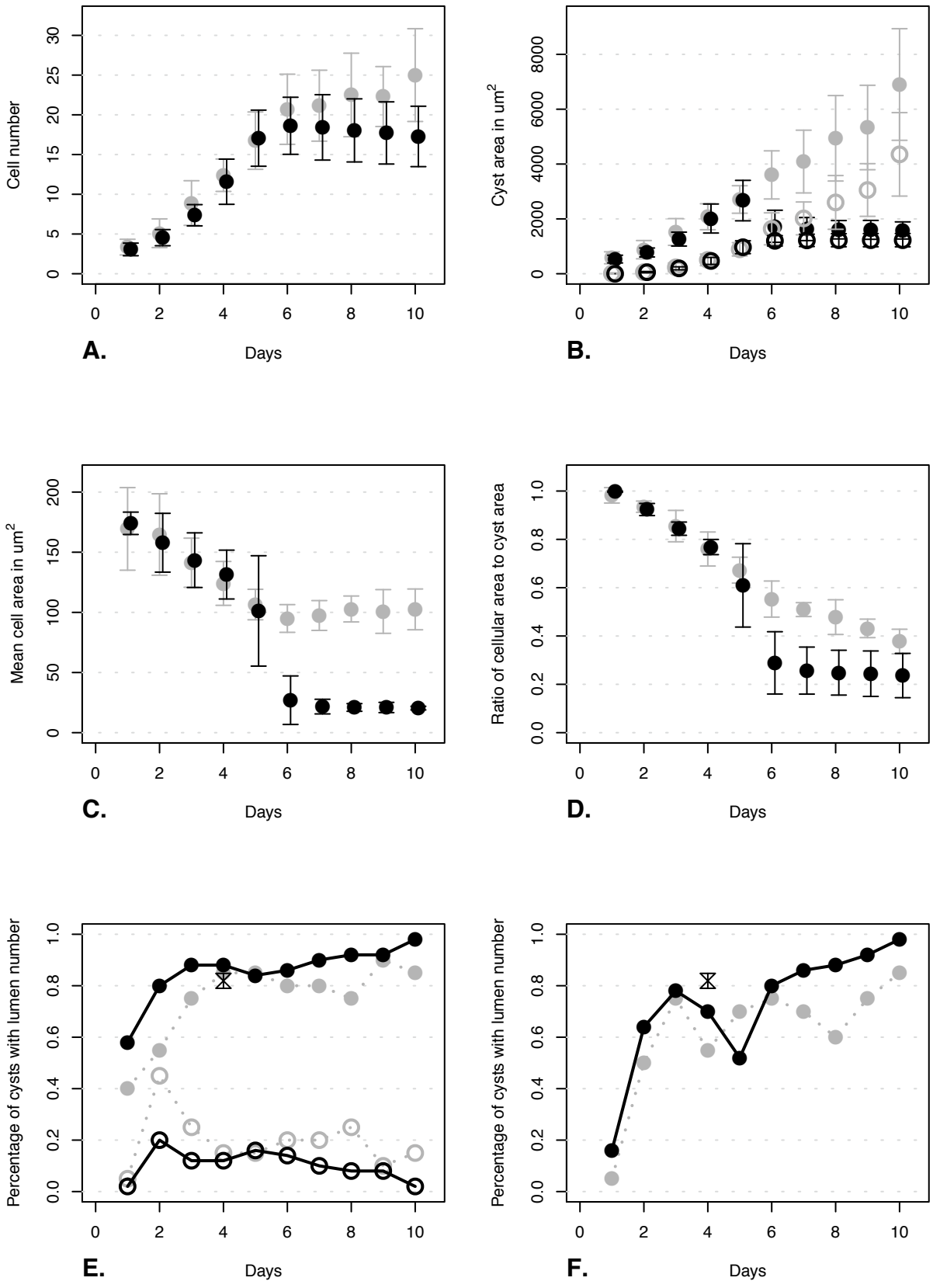
Appendix 1, Figure 12. Varied lambdaArea: la = 7.5



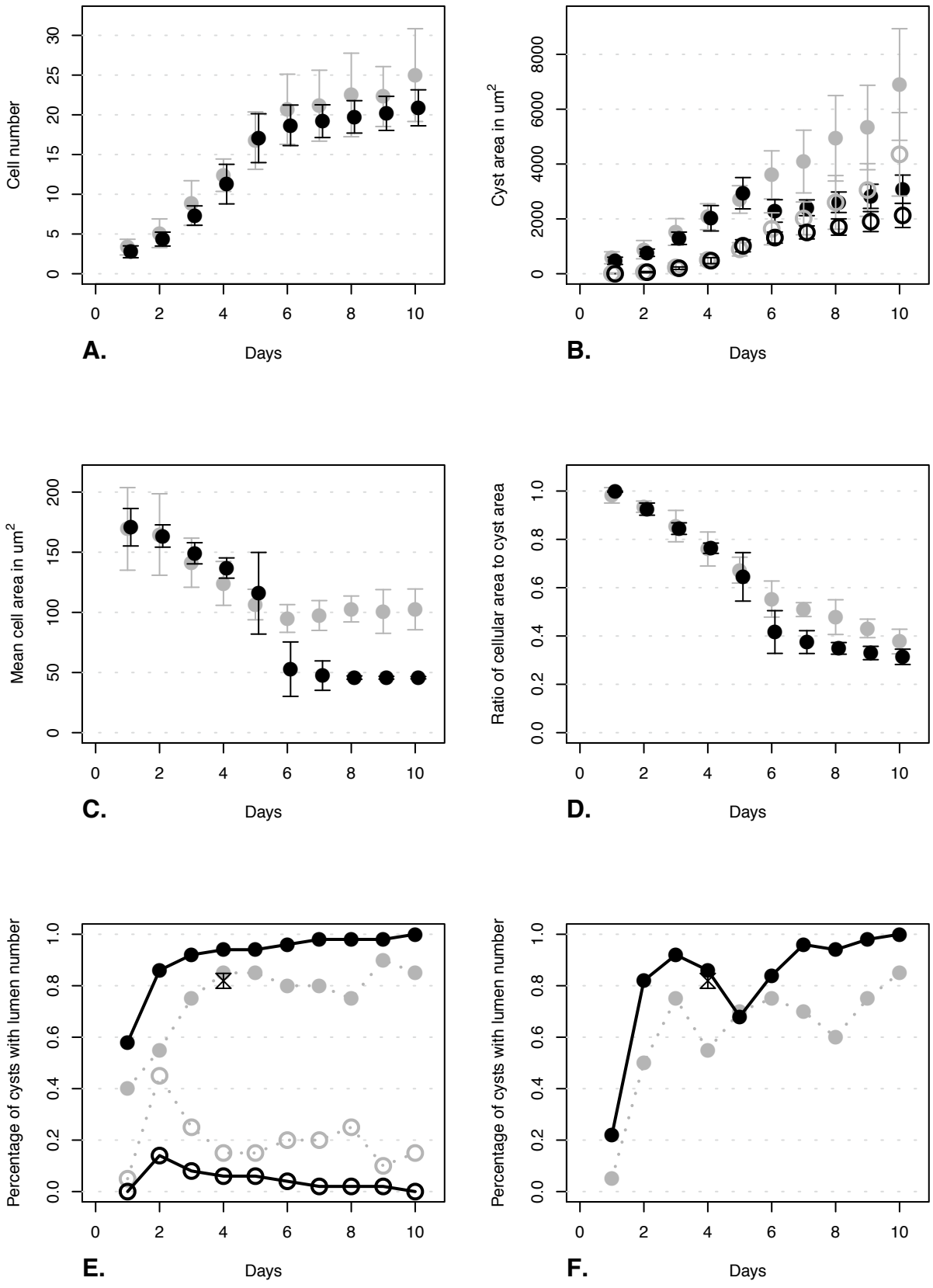
Appendix 1, Figure 13. Varied lambdaArea: la = 10



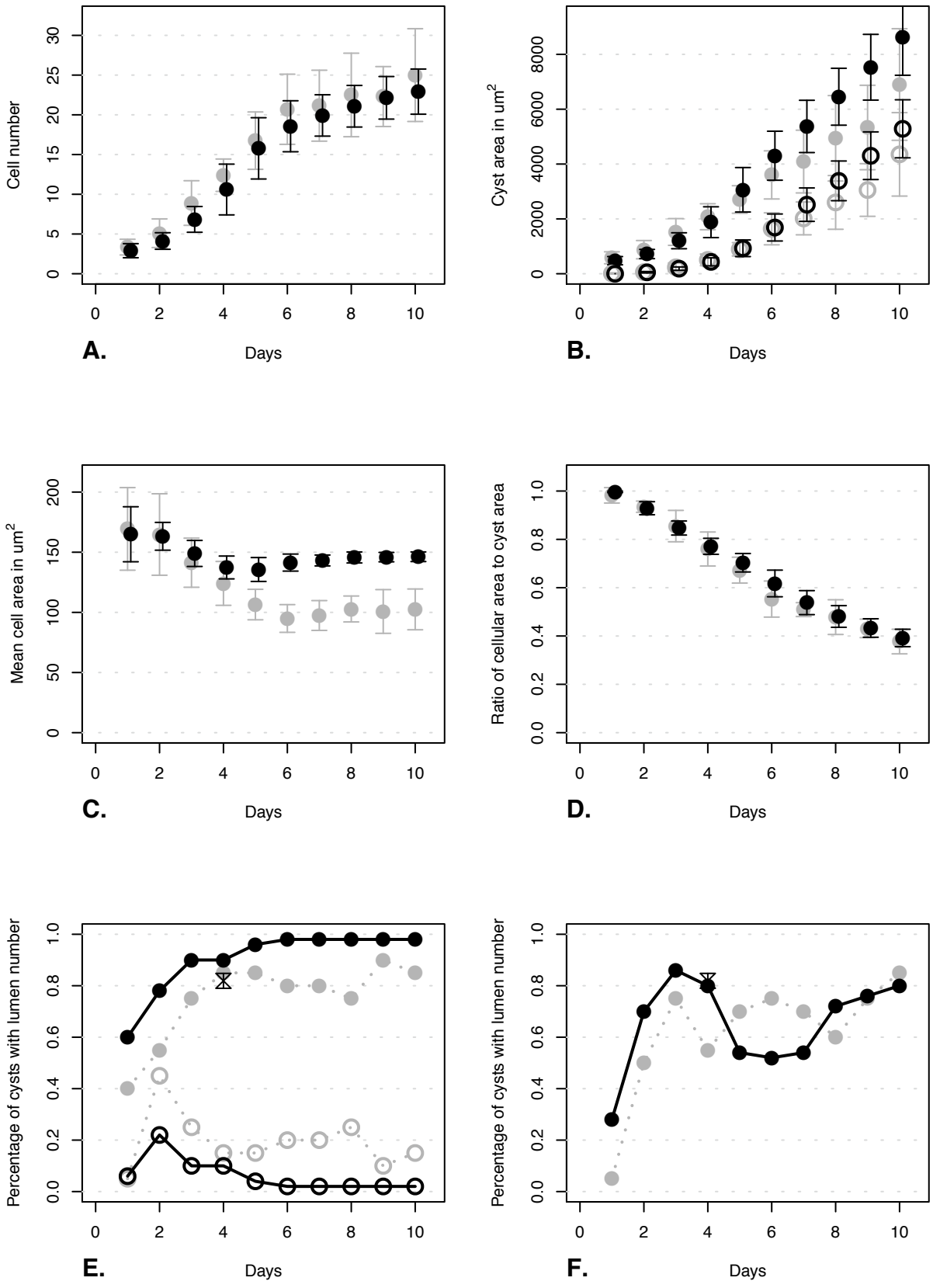
Appendix 1, Figure 14. Varied stableTargetArea: sta = 12



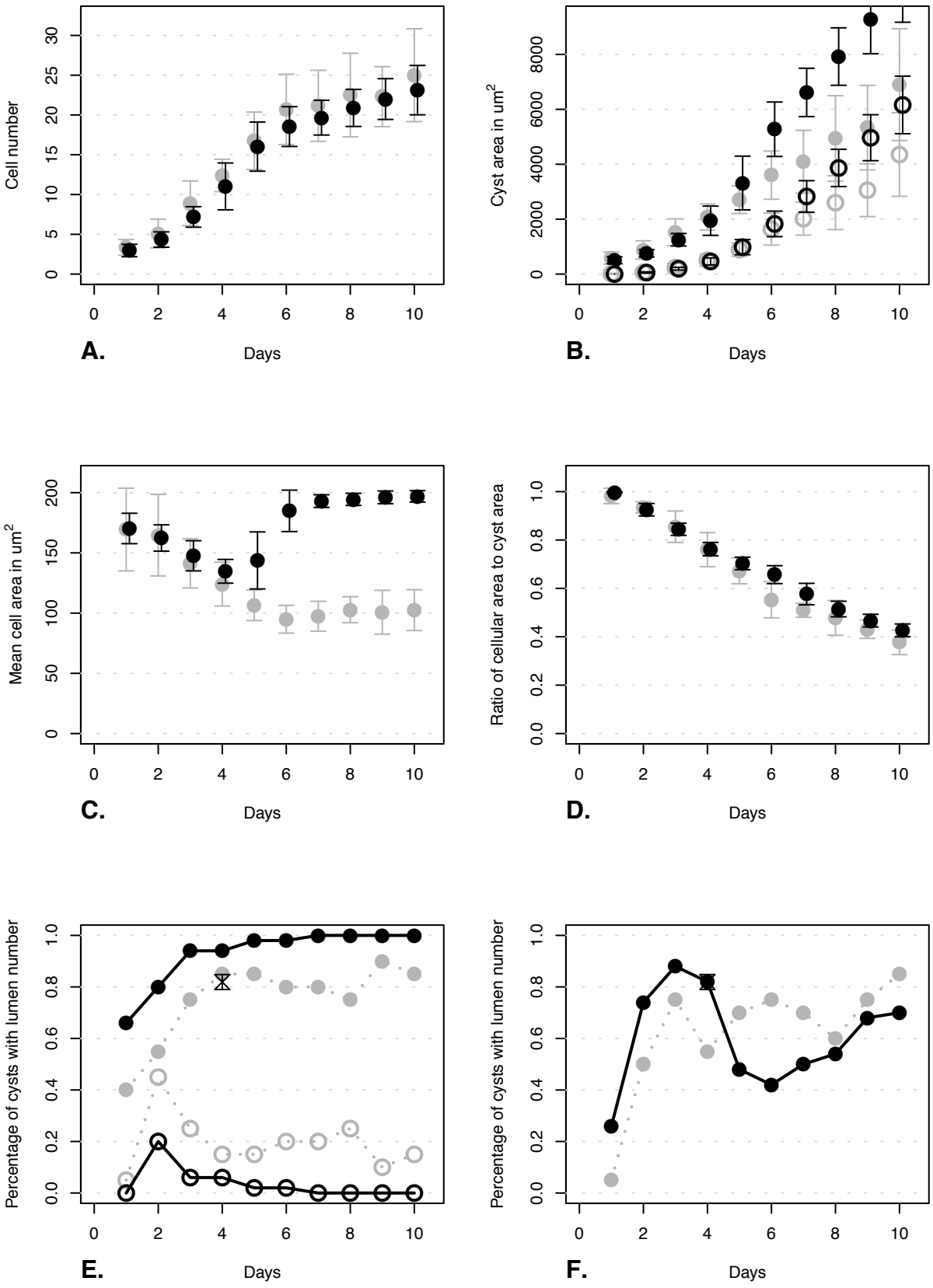
Appendix 1, Figure 15. Varied stableTargetArea: sta = 24



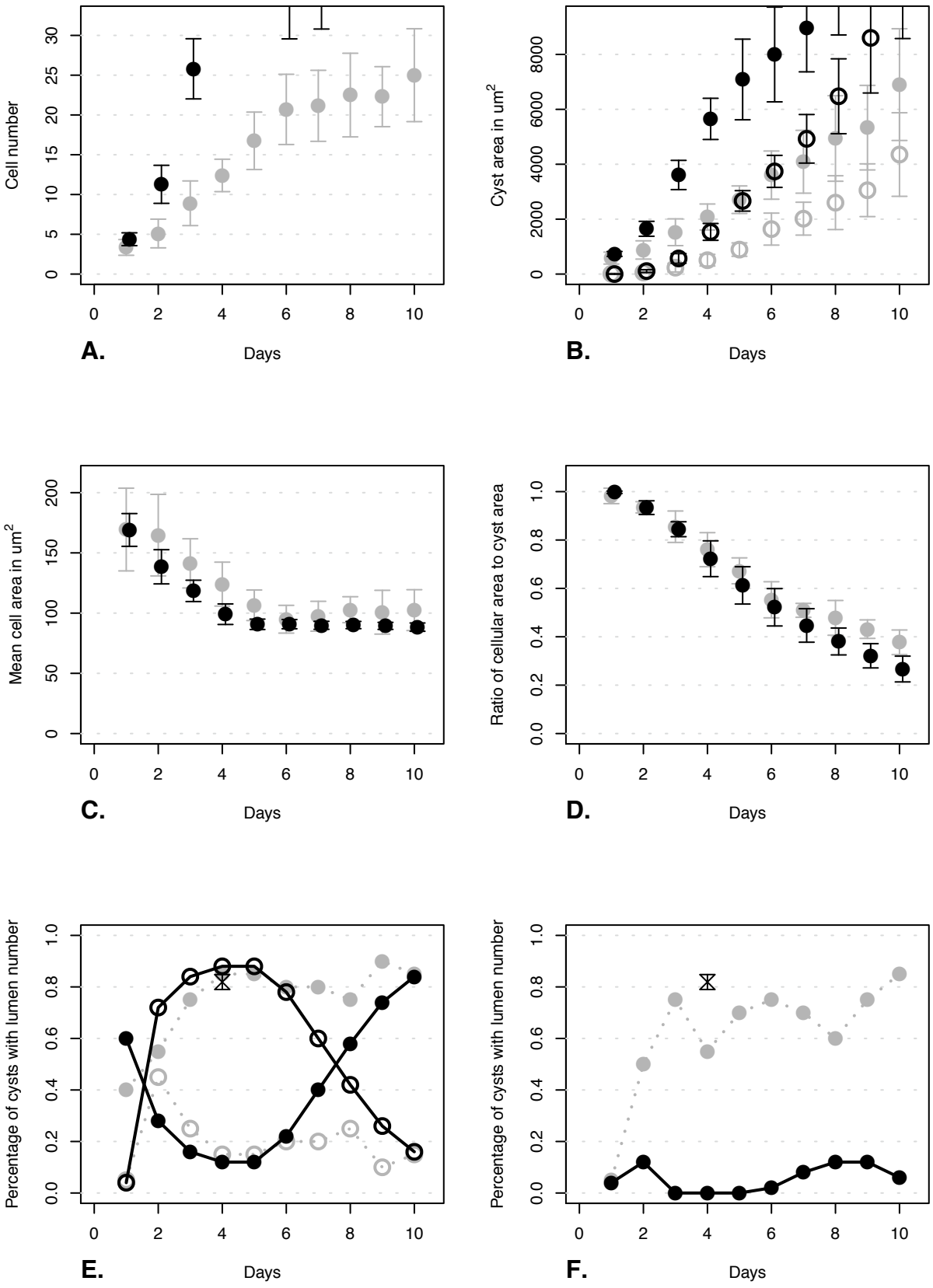
Appendix 1, Figure 16. Varied stableTargetArea: sta = 72



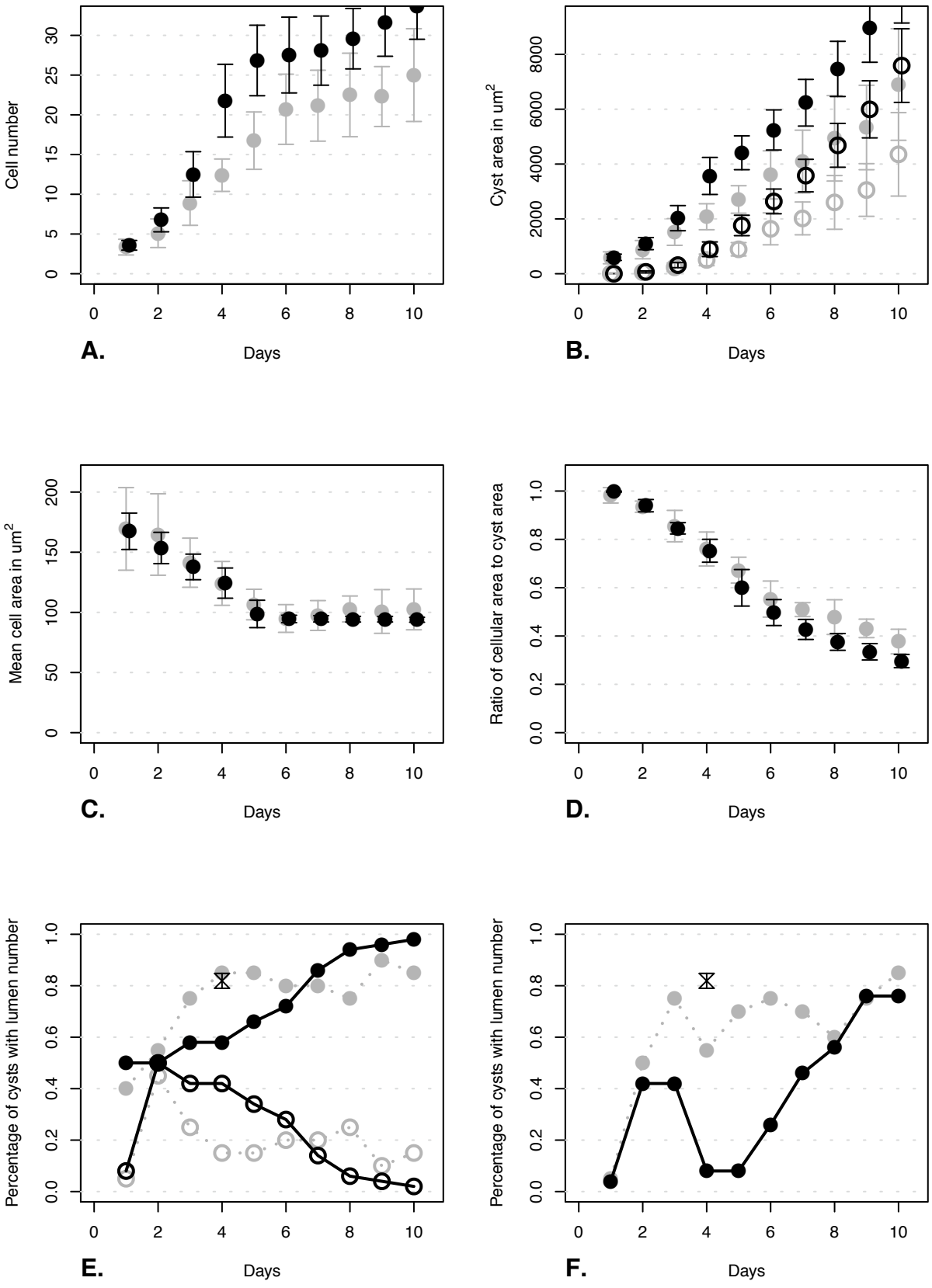
Appendix 1, Figure 17. Varied stableTargetArea sta = 96



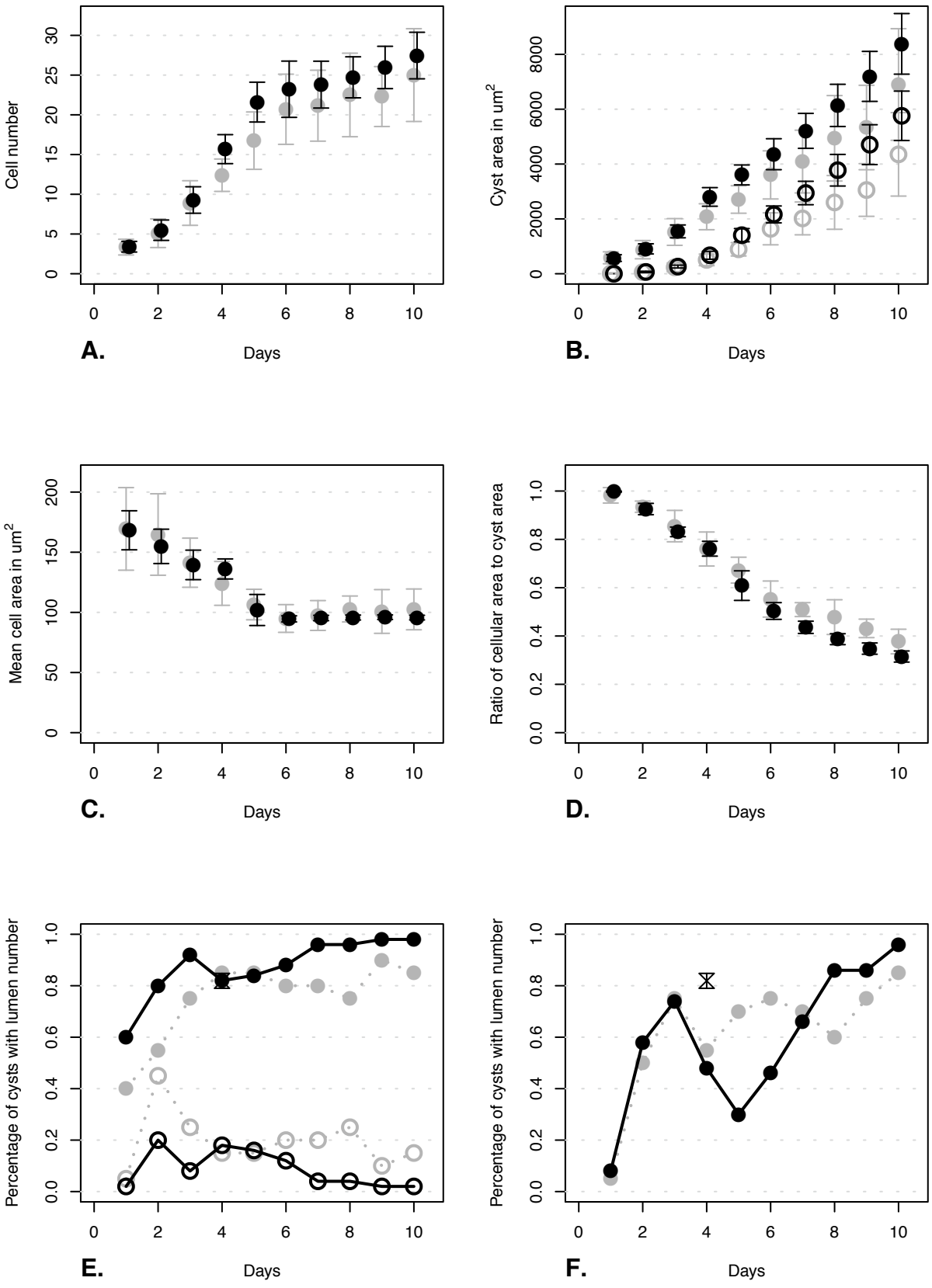
Appendix 1, Figure 18. Varied cellCycle: cc = 35



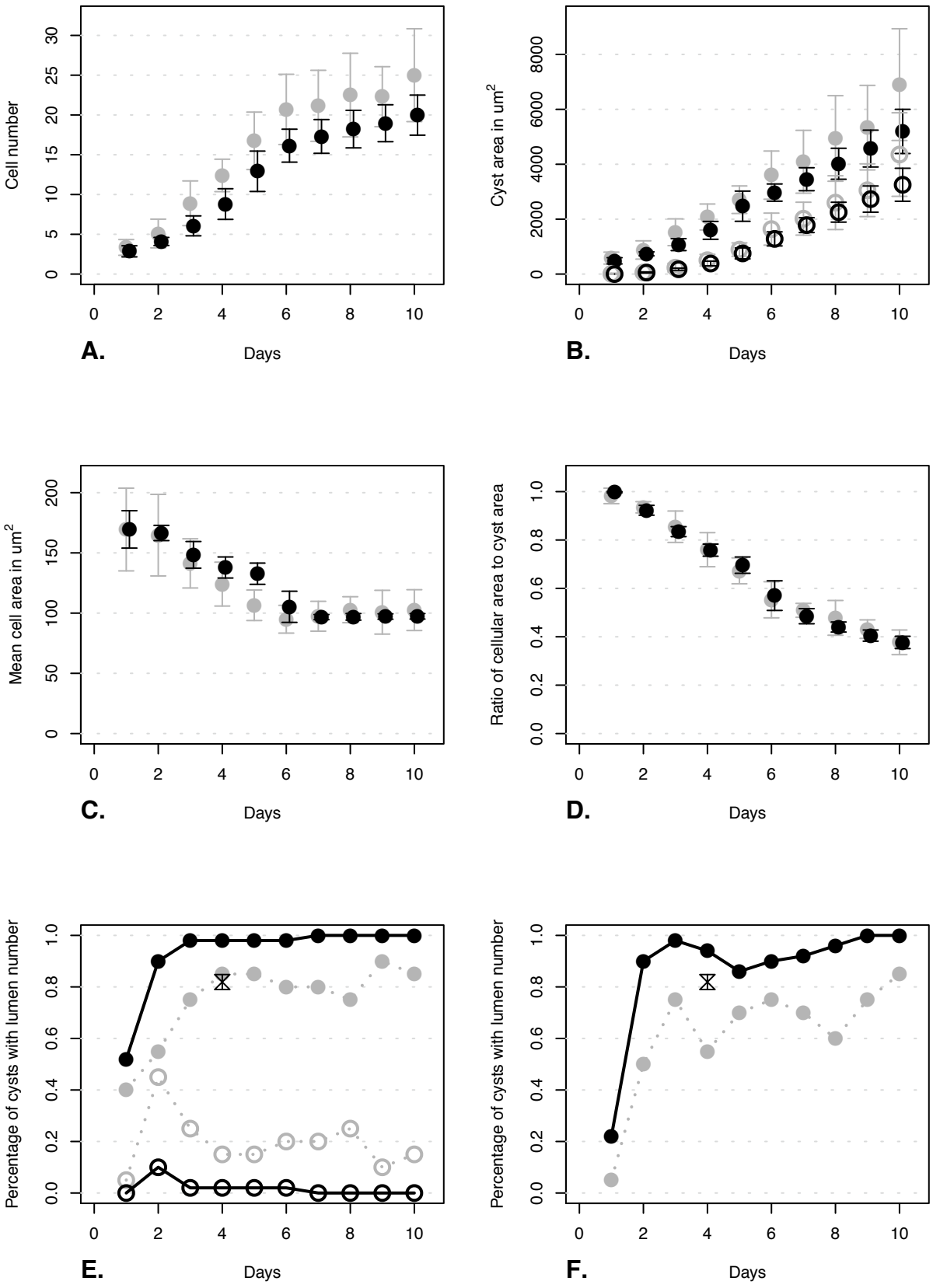
Appendix 1, Figure 19. Varied cellCycle: cc = 50



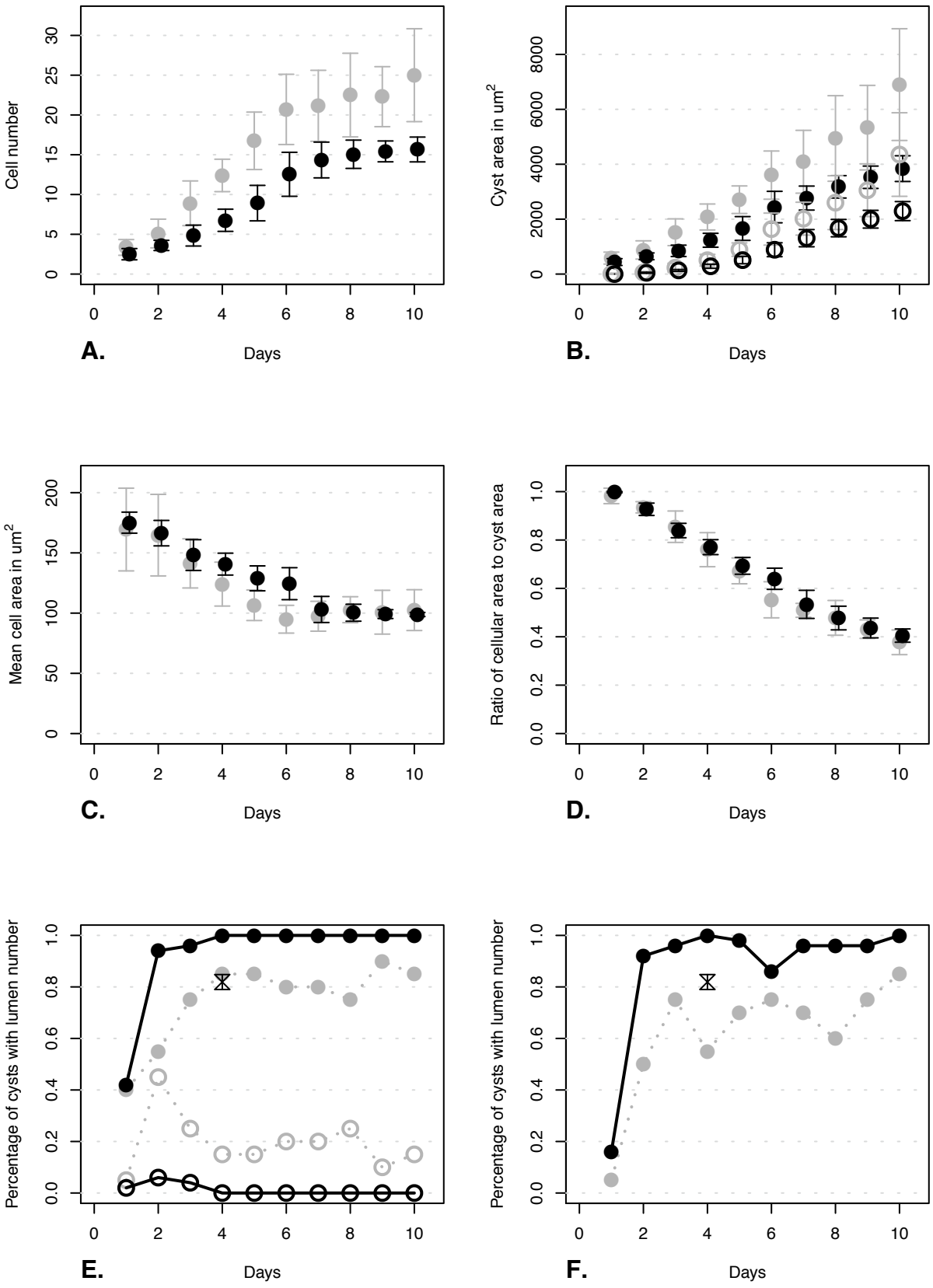
Appendix 1, Figure 20. Varied cellCycle: cc = 60



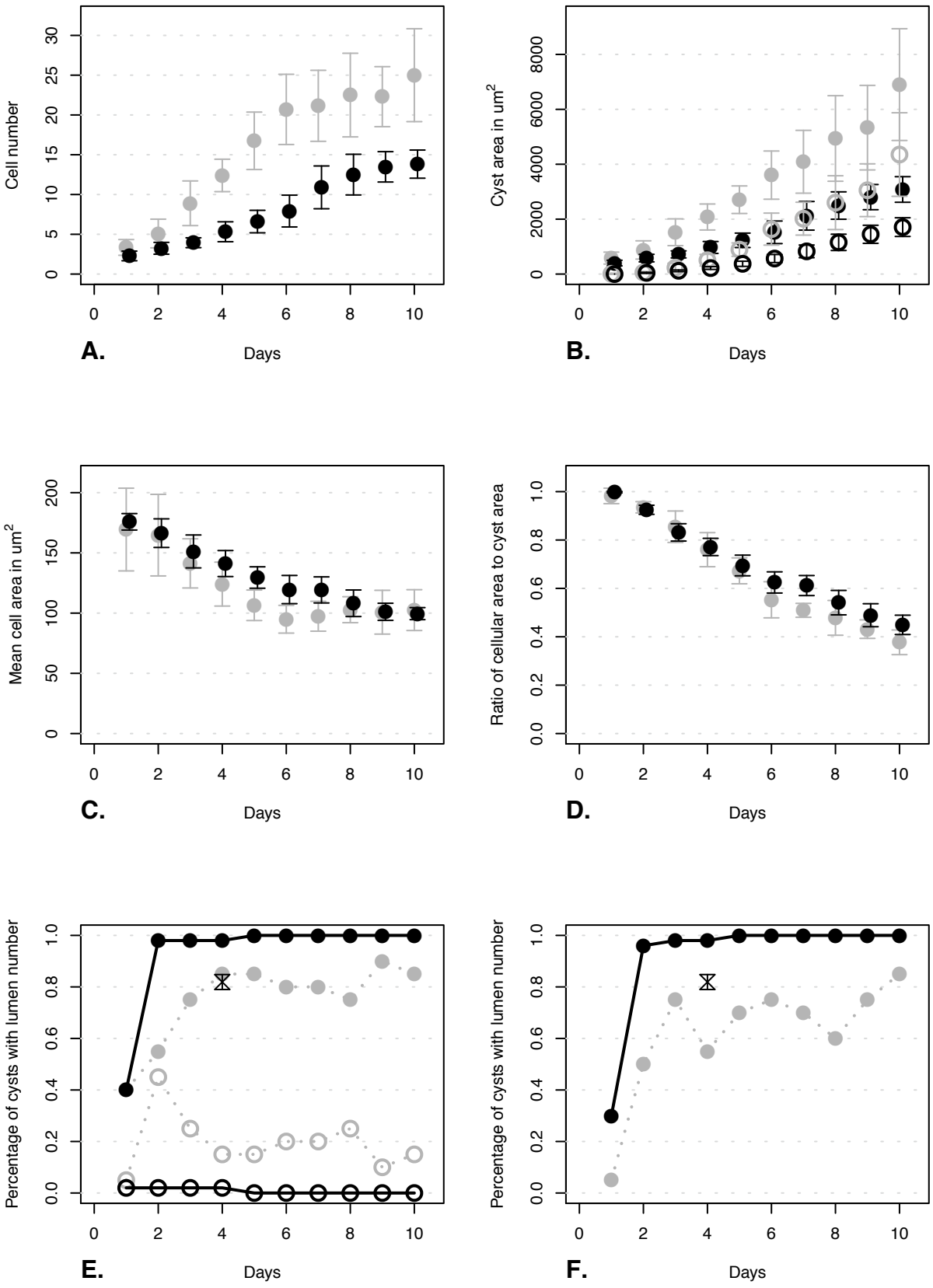
Appendix 1, Figure 21. Varied cellCycle: cc = 80



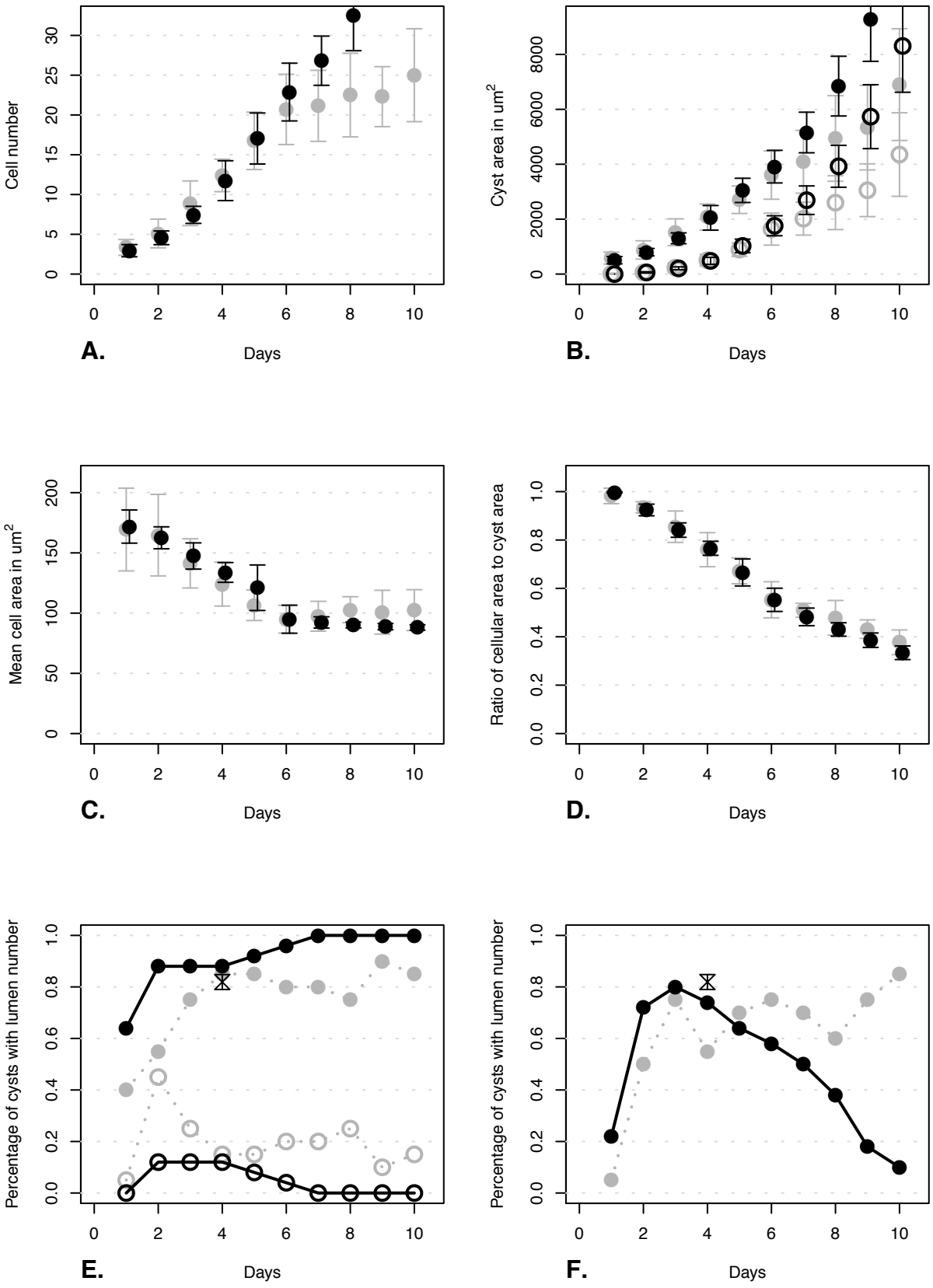
Appendix 1, Figure 22. Varied cellCycle: cc = 100



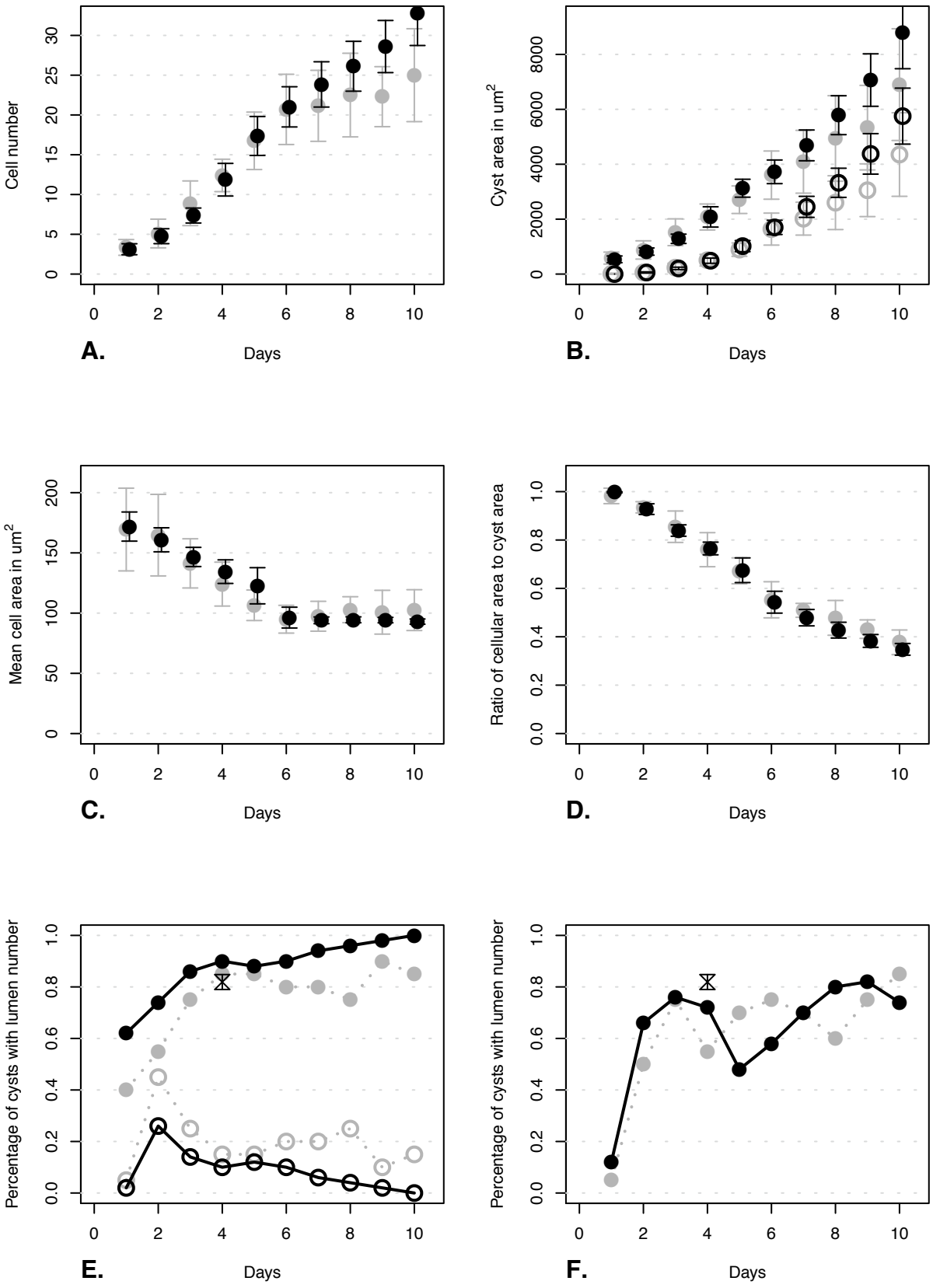
Appendix 1, Figure 23. Varied cellCycle: cc = 140



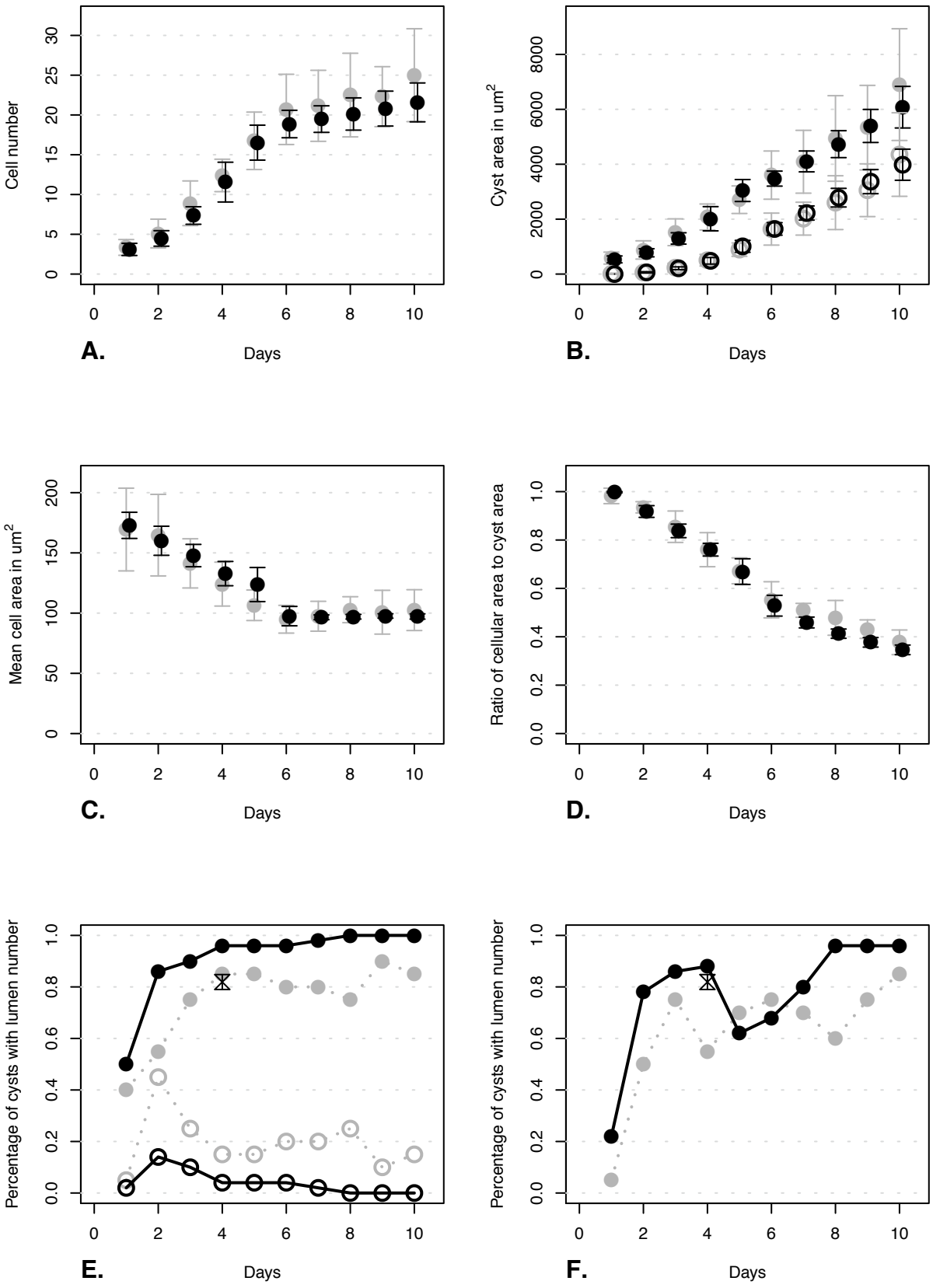
Appendix 1, Figure 24. Varied stableCycleDelay: scd = 0.5



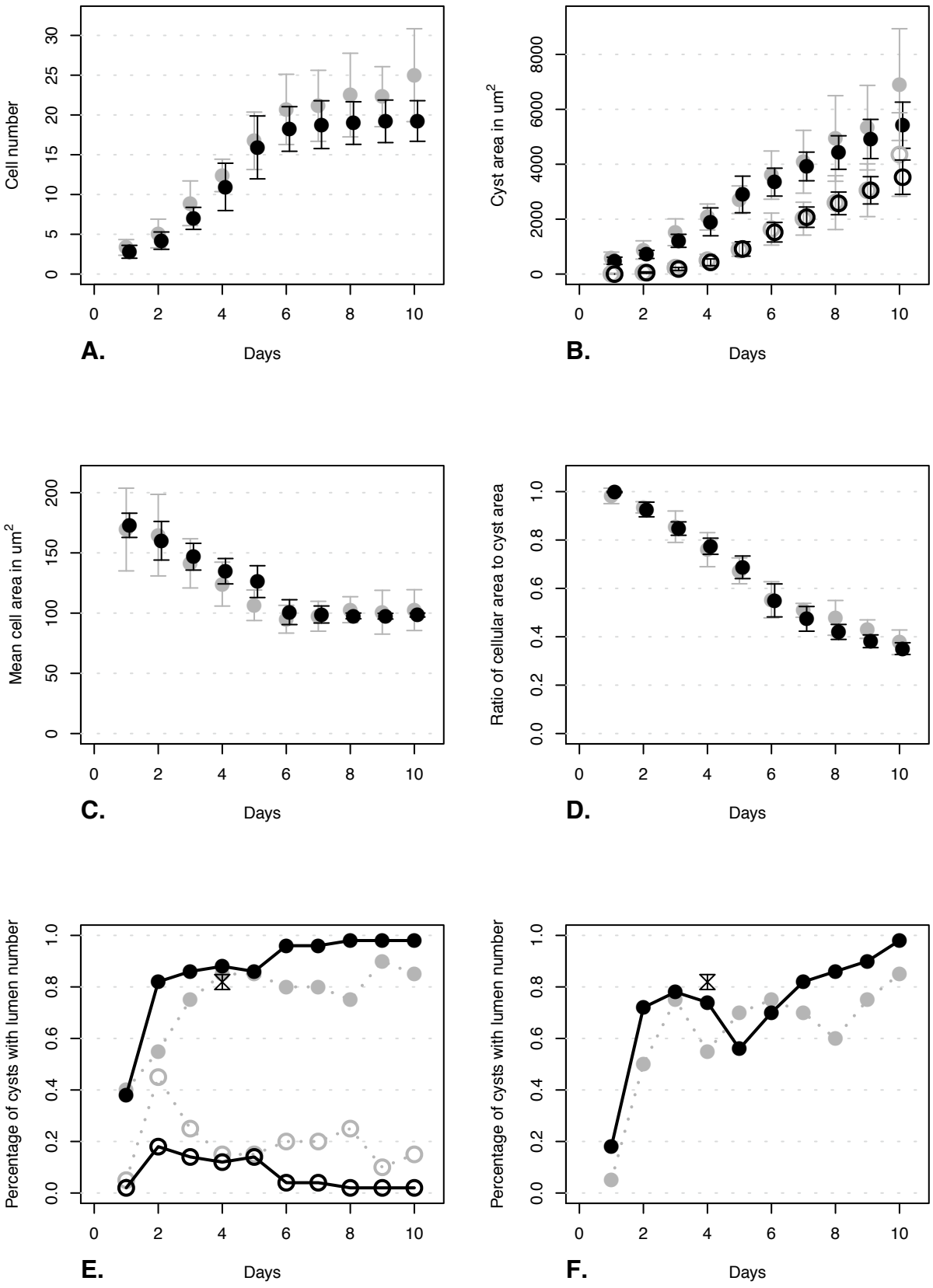
Appendix 1, Figure 25. Varied stableCycleDelay: scd = 0.7



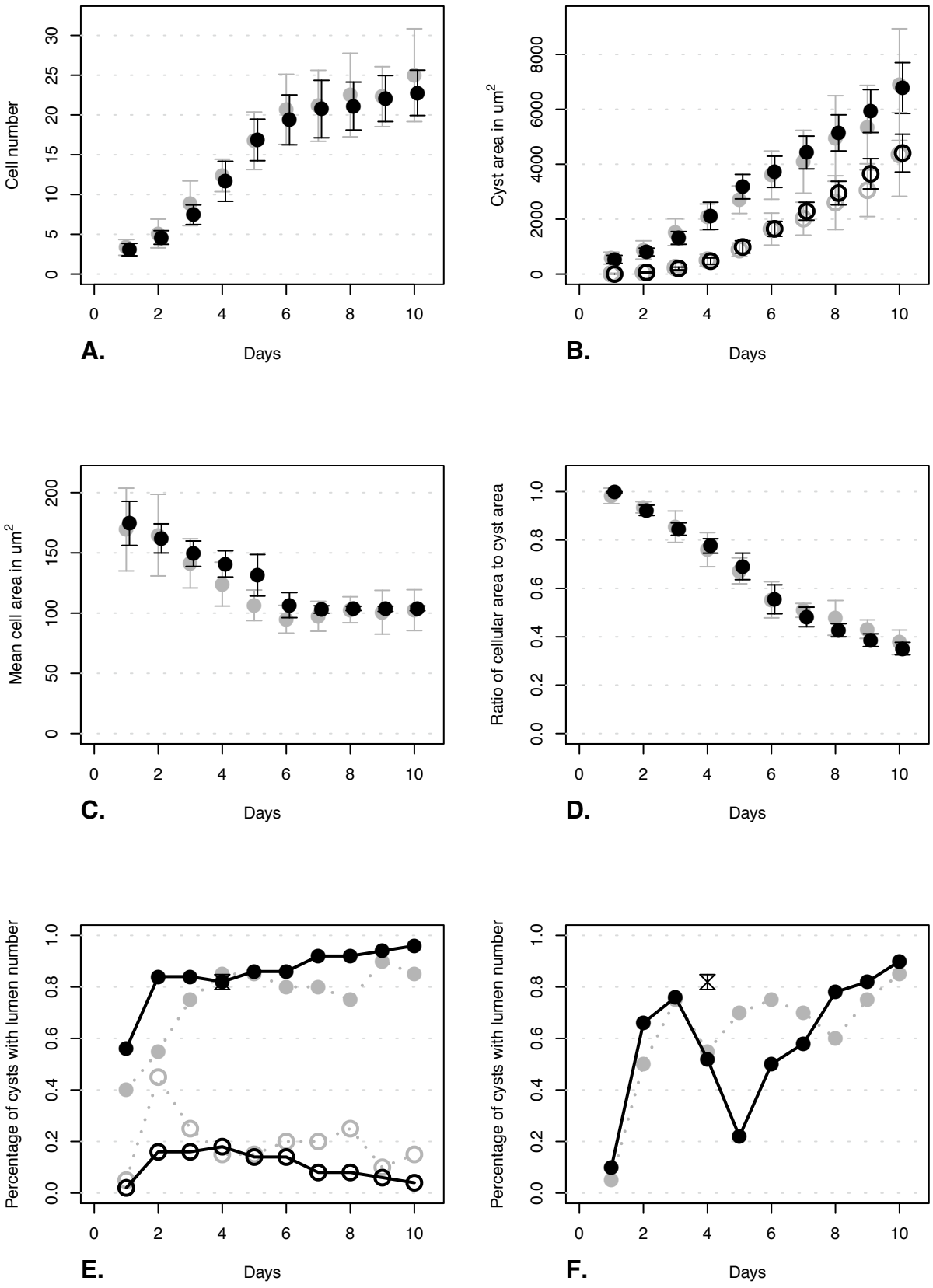
Appendix 1, Figure 26. Varied stableCycleDelay: scd = 0.9



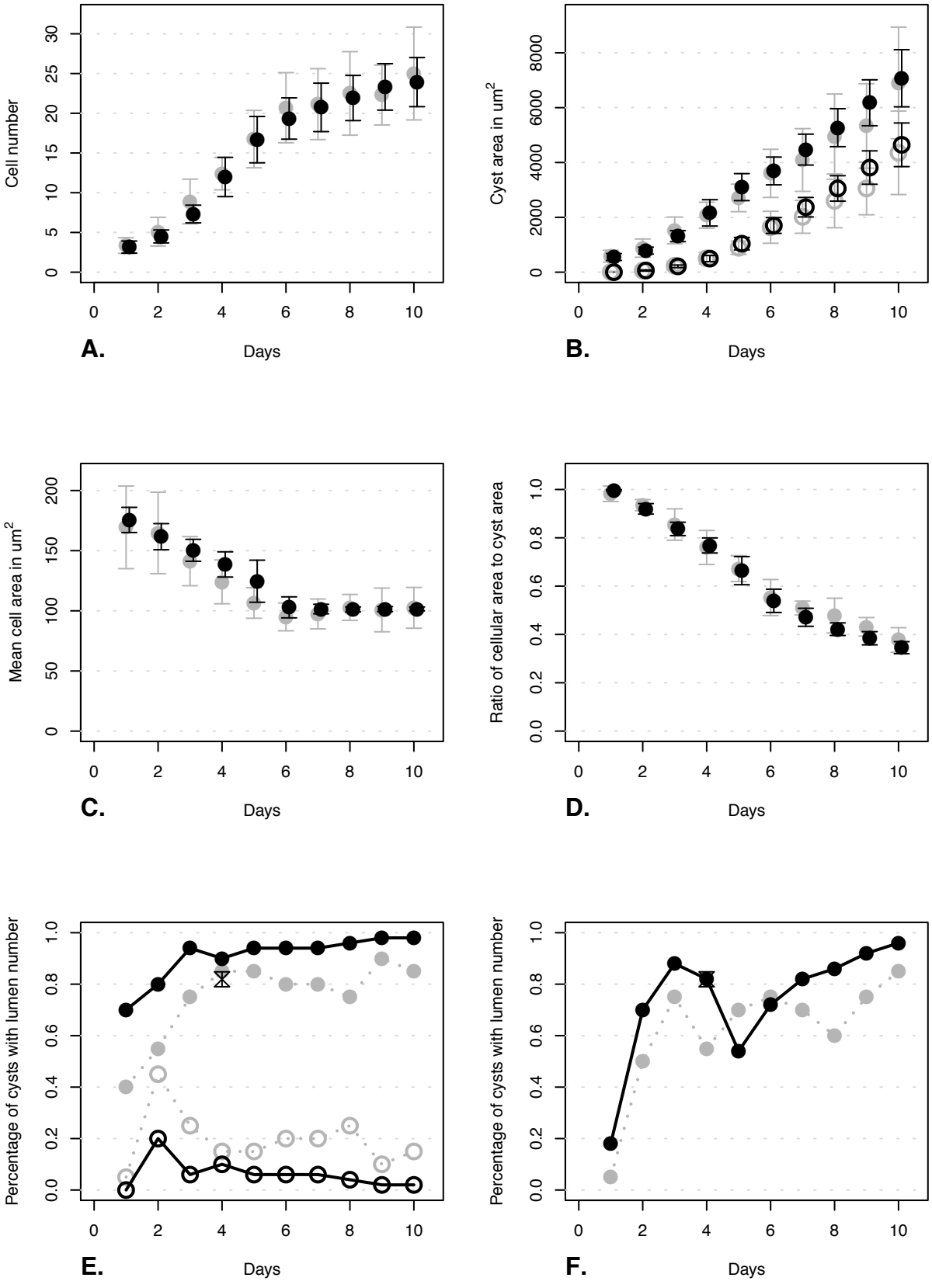
Appendix 1, Figure 27. Varied stableCycleDelay: scd = 0.95



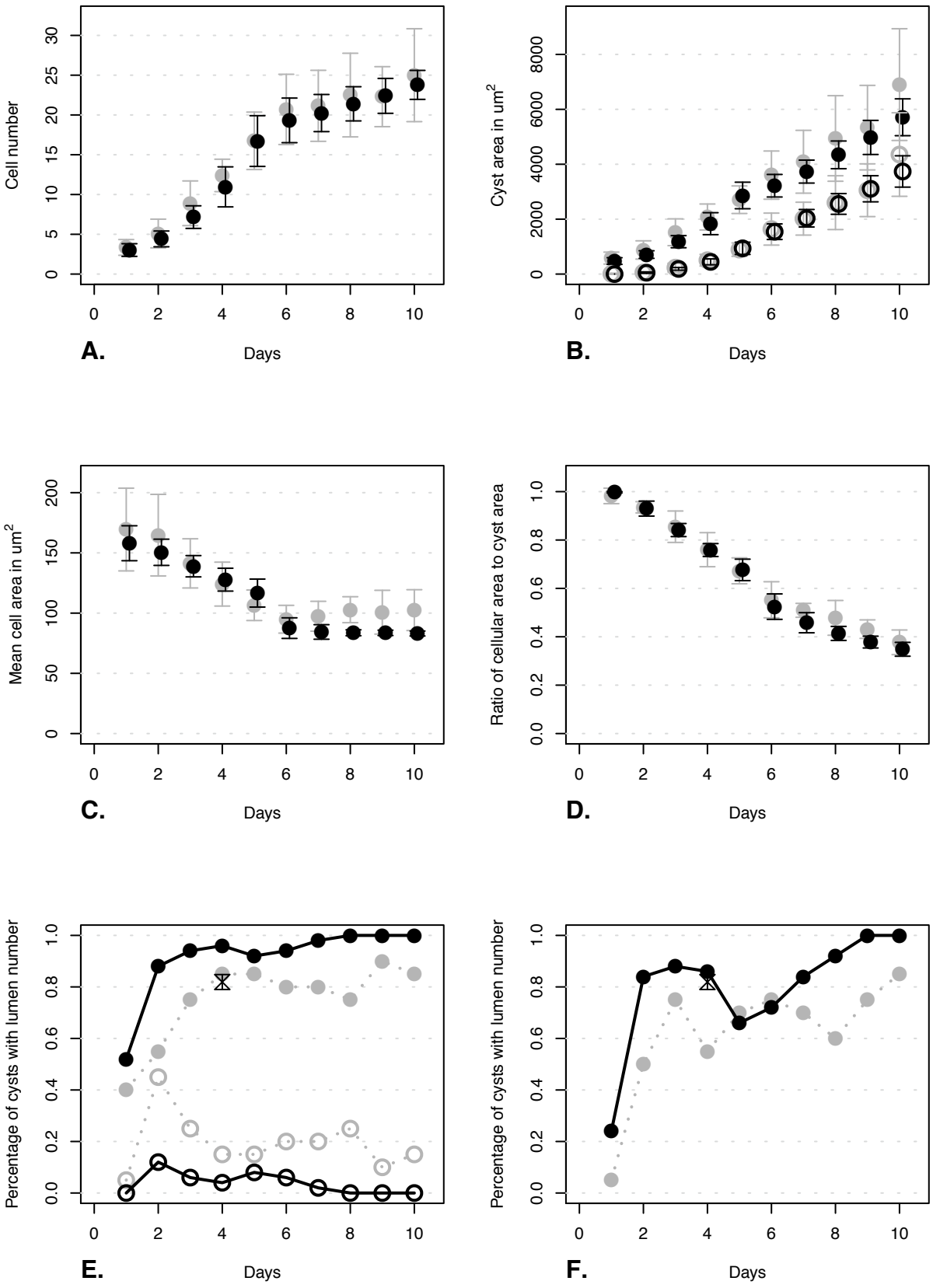
Appendix 1, Figure 28. Varied lambdaPerim: lp = .5



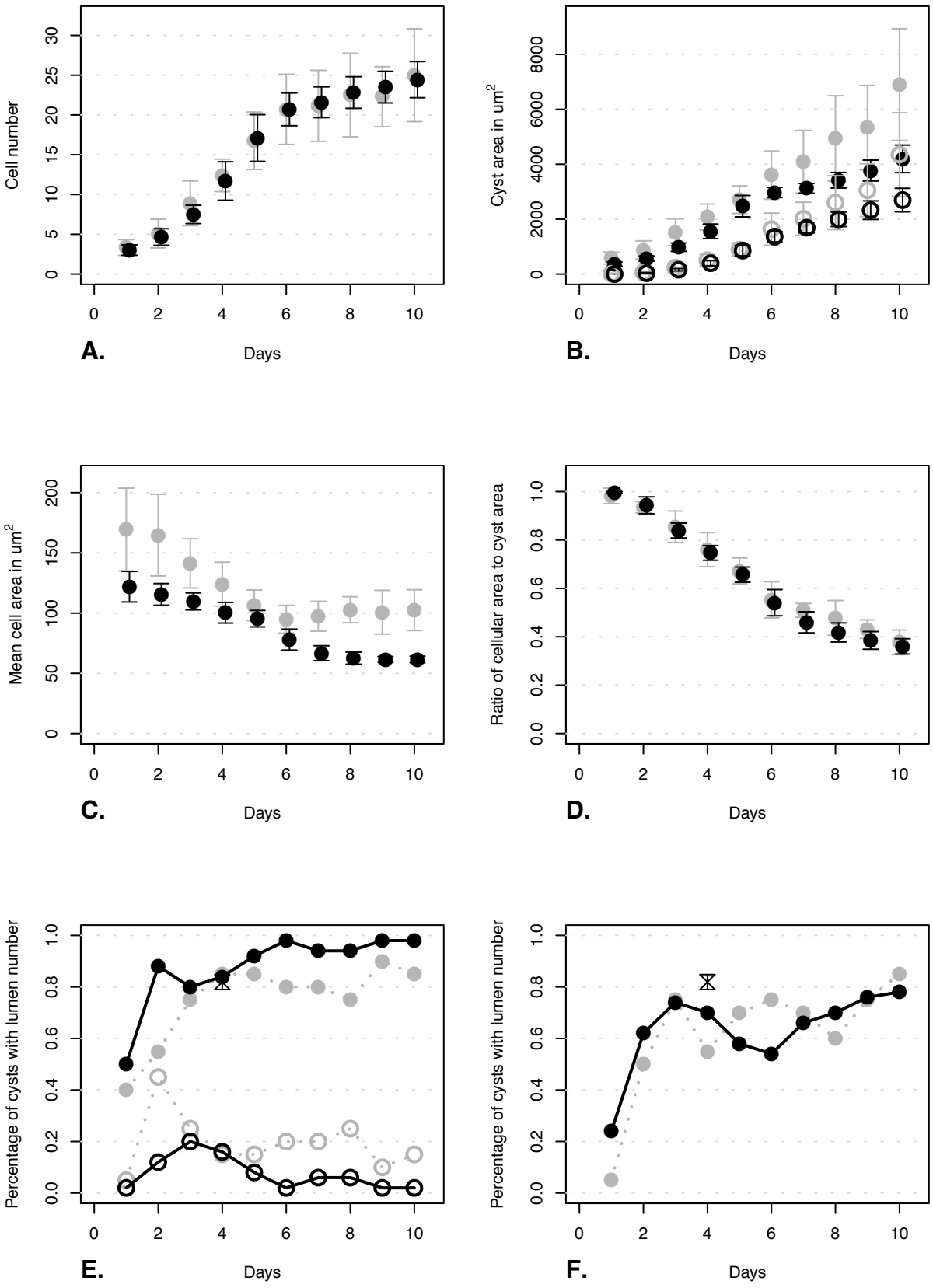
Appendix 1, Figure 29. Varied lambdaPerim: lp = 1.25



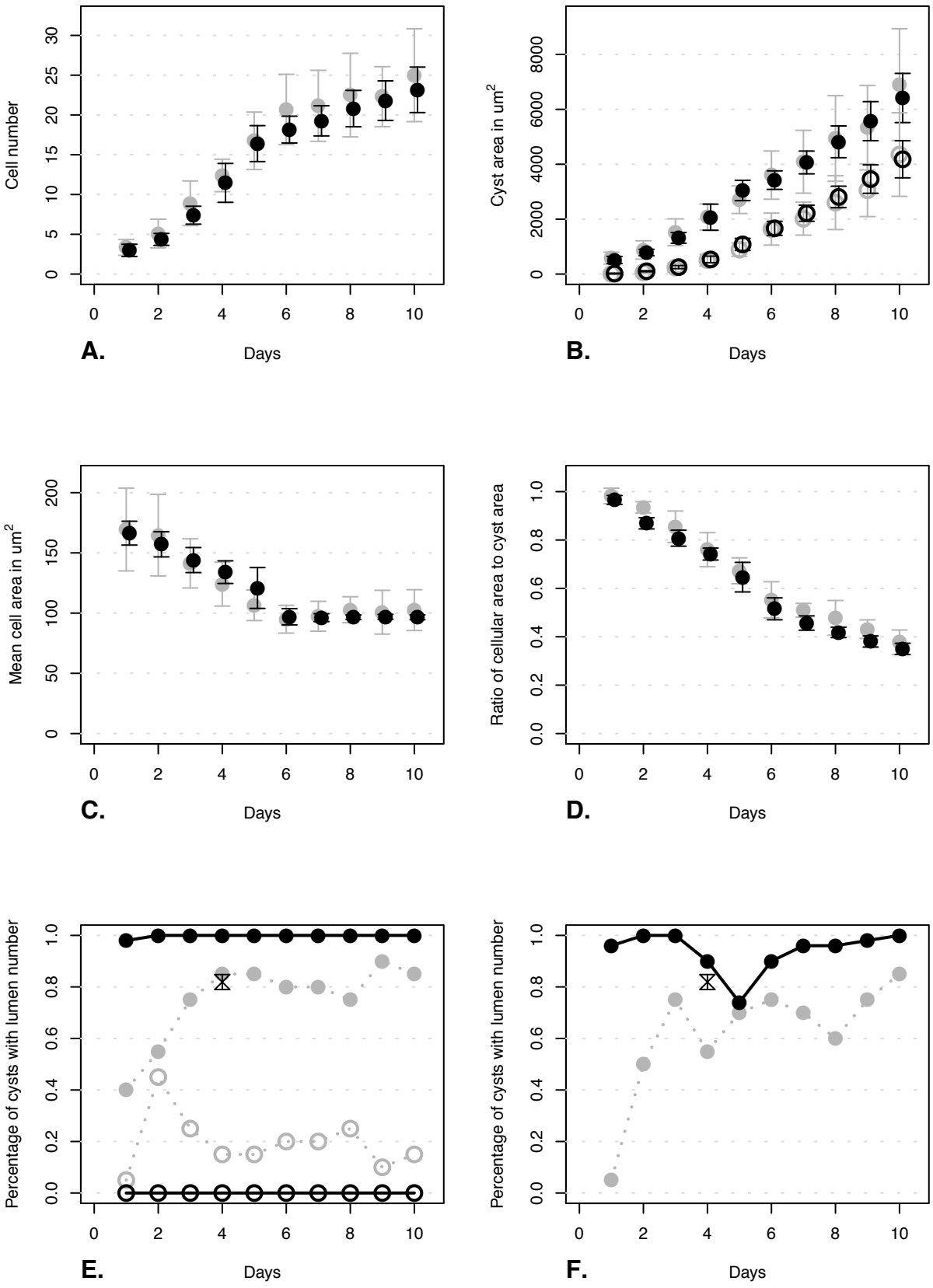
Appendix 1, Figure 30. Varied lambdaPerim: lp = 5



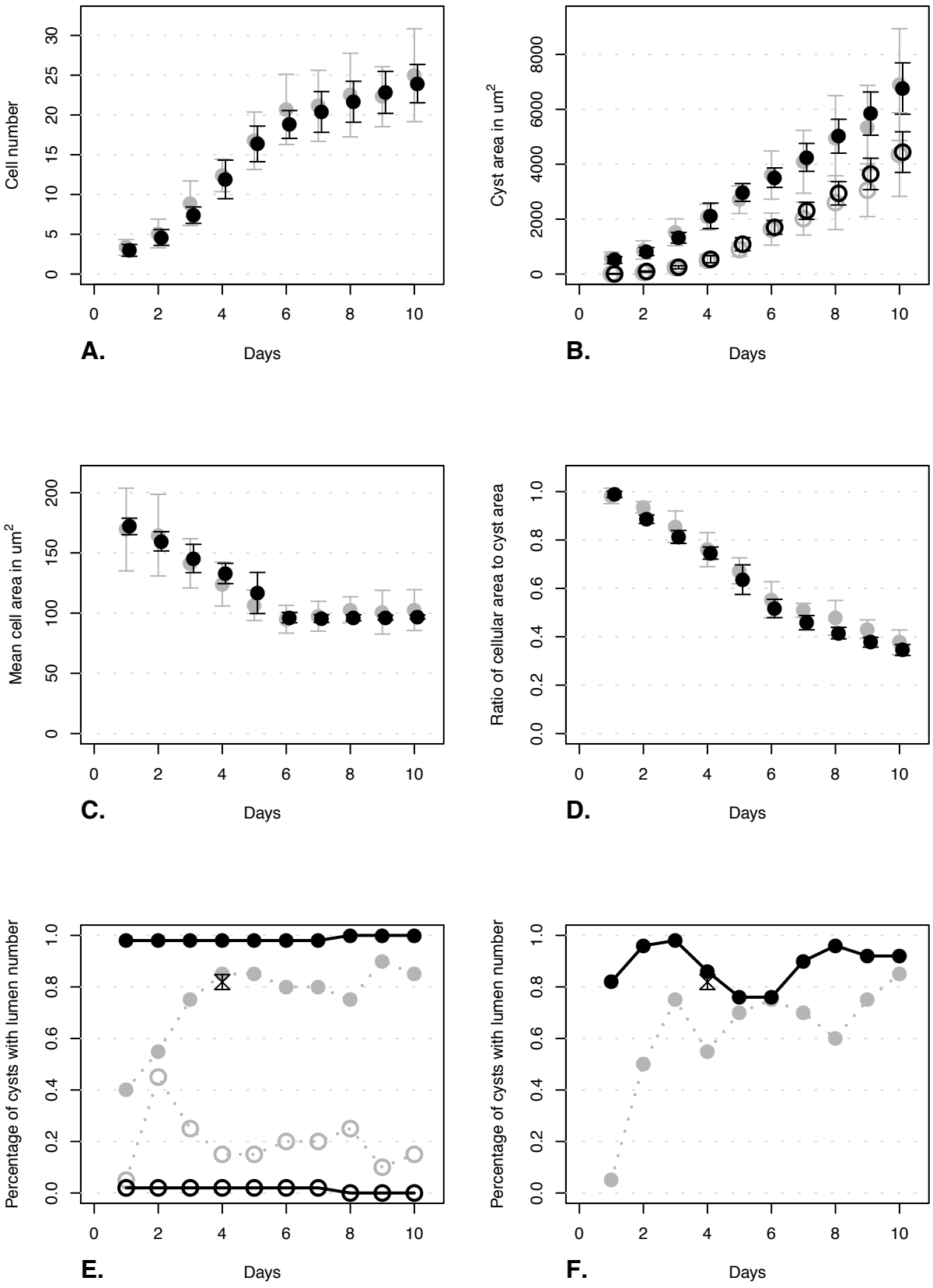
Appendix 1, Figure 31. Varied lambdaPerim: lp = 10



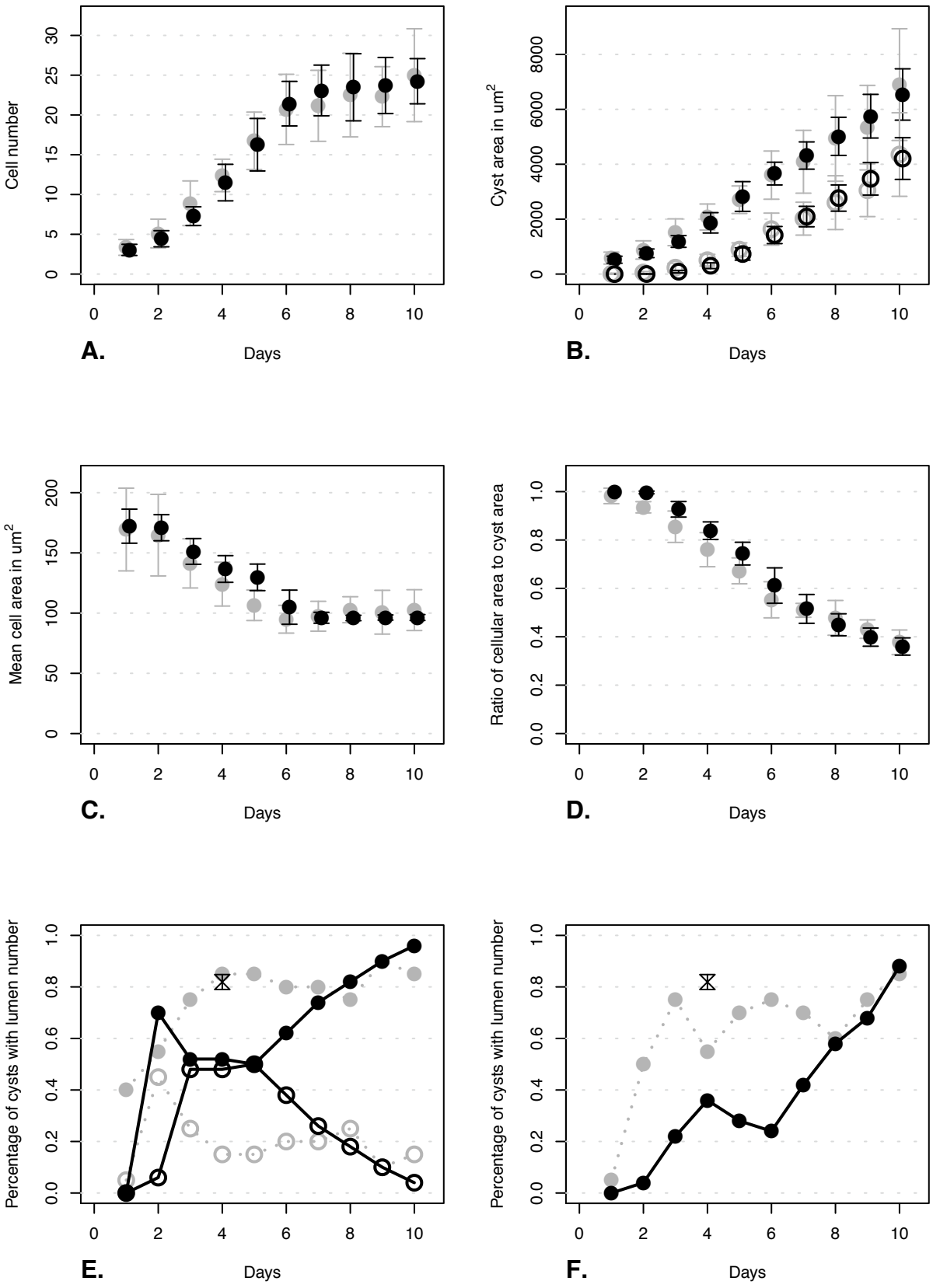
Appendix 1, Figure 32. Varied polarDelay: pd = 10



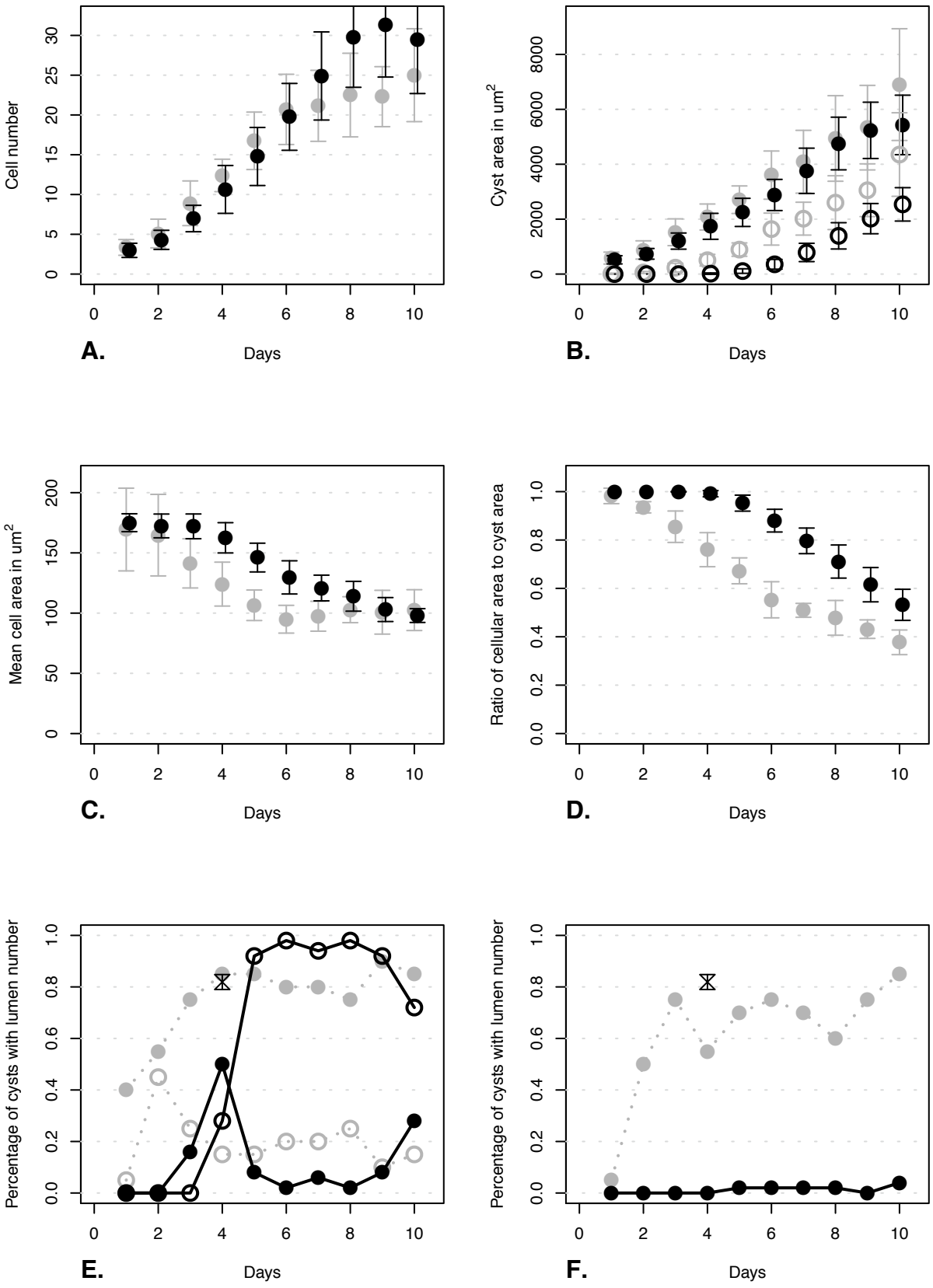
Appendix 1, Figure 33. Varied polarDelay: pd = 21



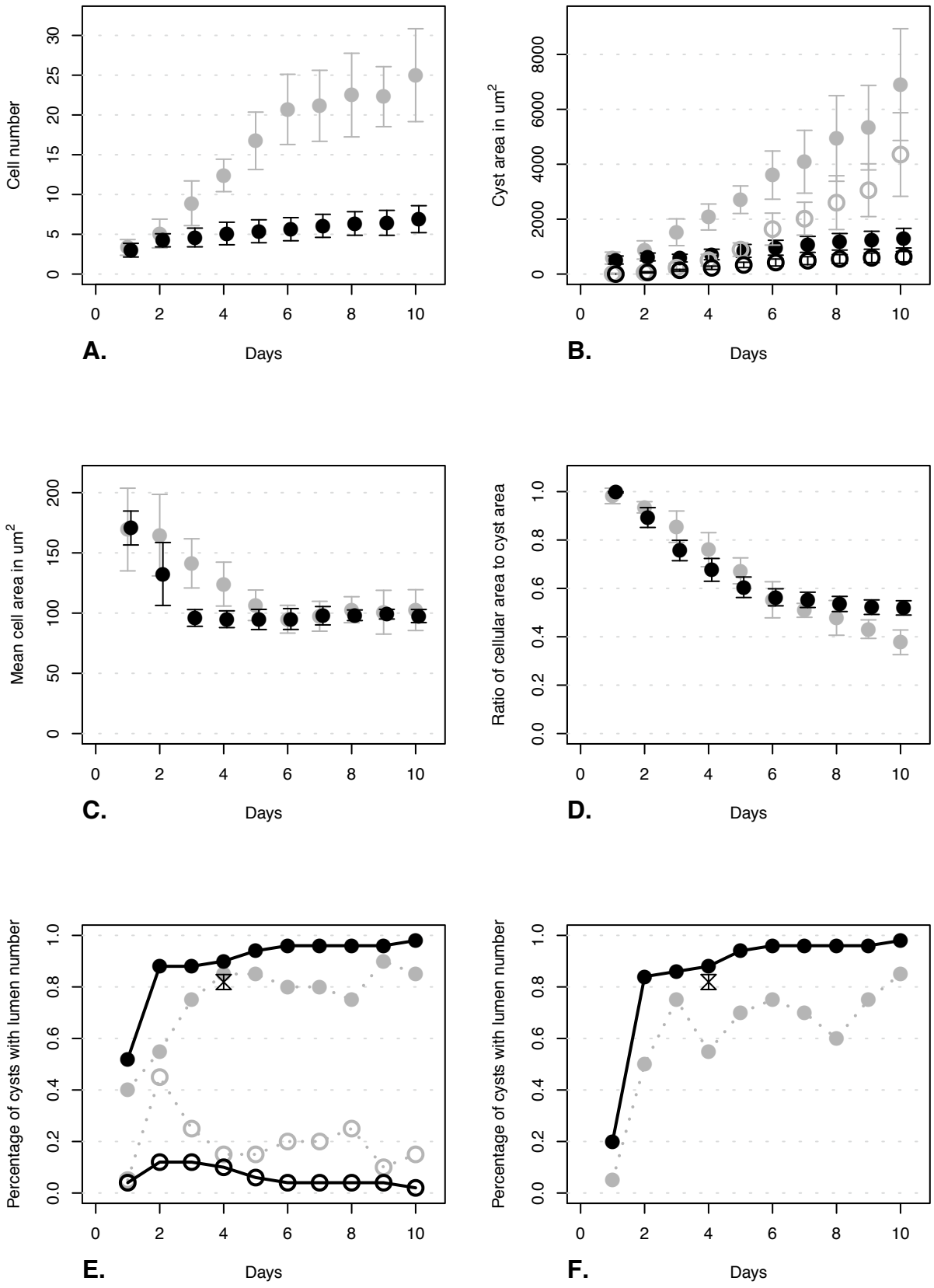
Appendix 1, Figure 34. Varied polarDelay: pd = 84



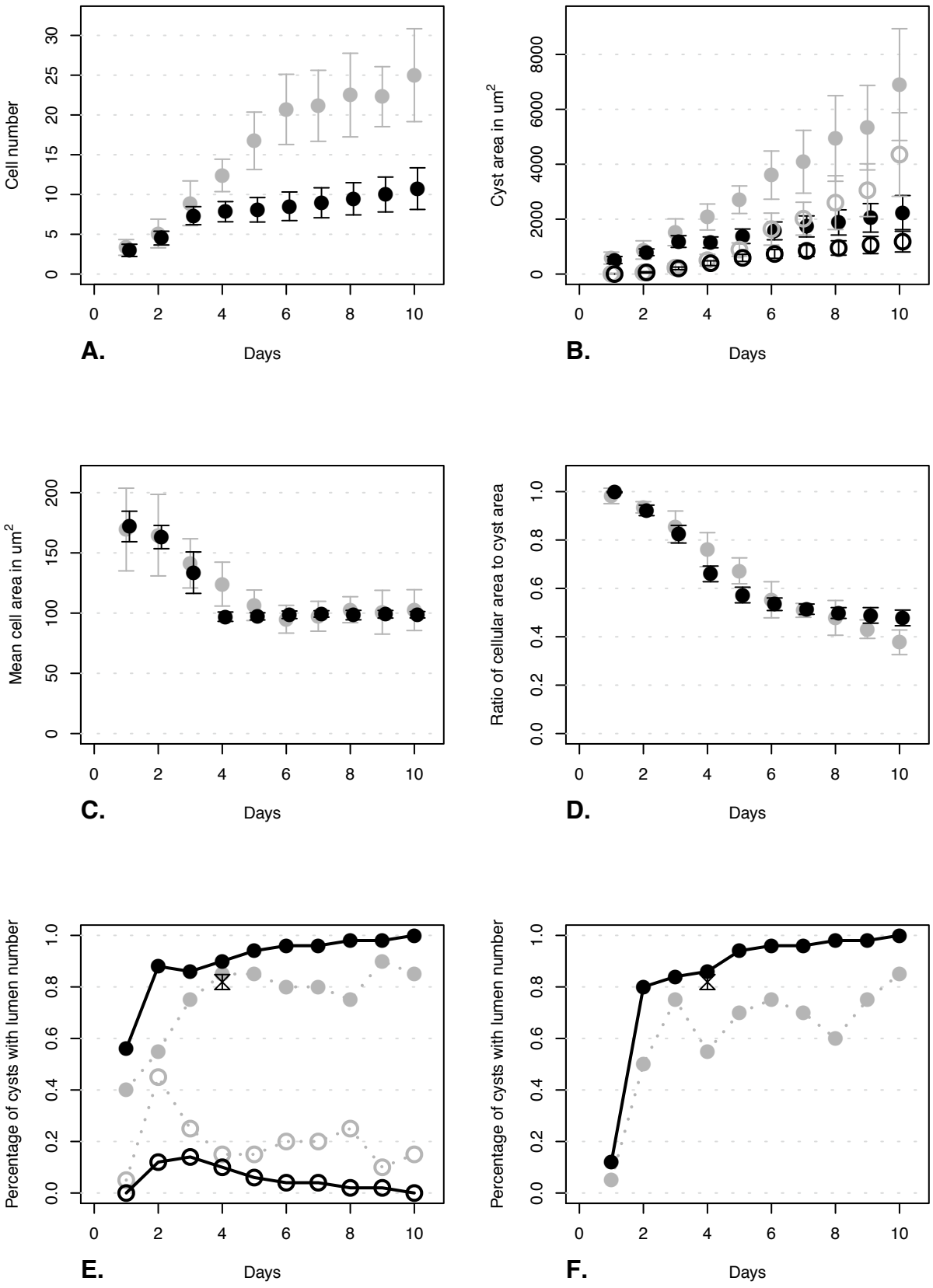
Appendix 1, Figure 35. Varied polarDelay: pd = 200



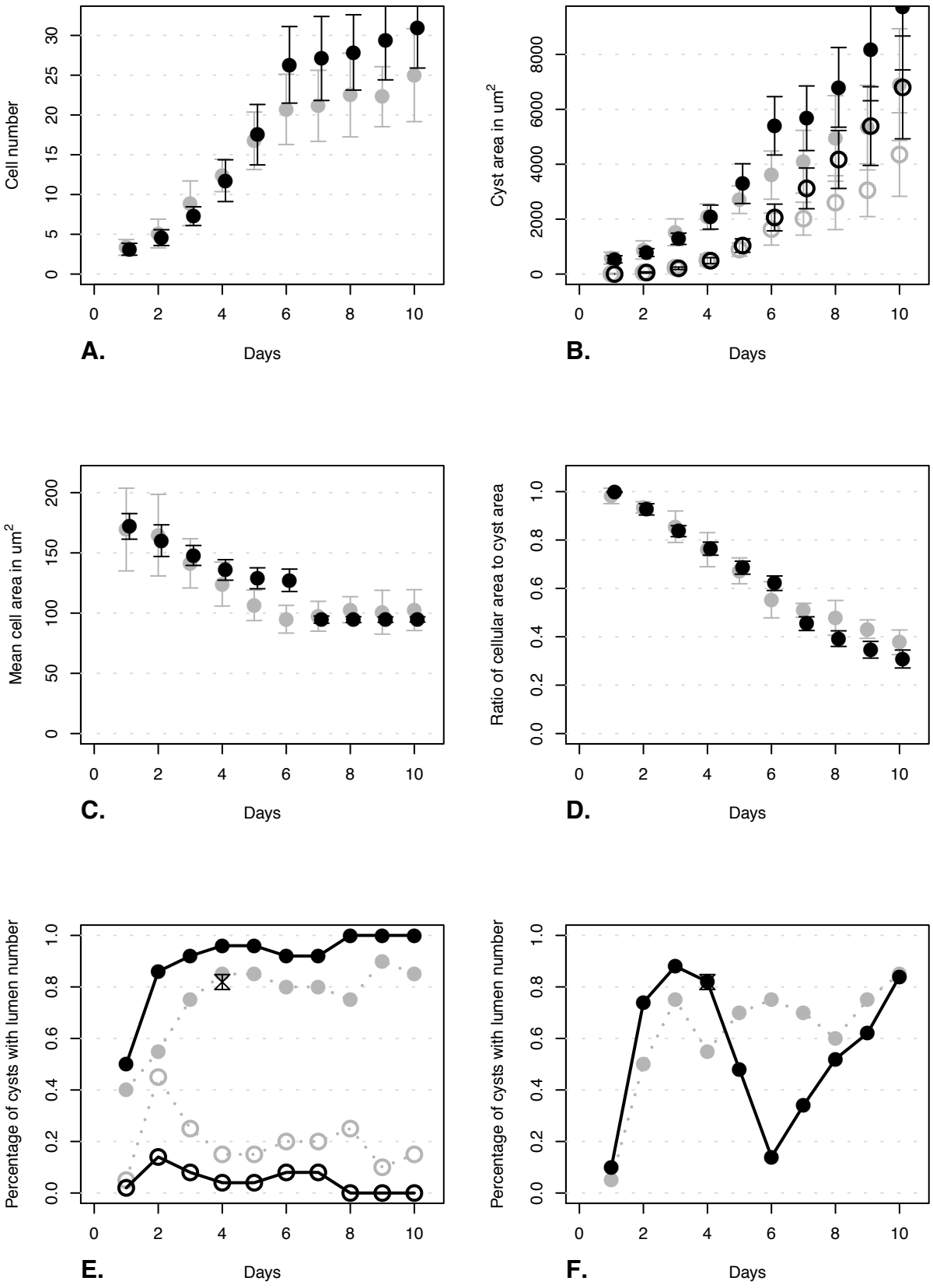
Appendix 1, Figure 36. Varied shiftDelay with stableRatio = 100000: sd = 50



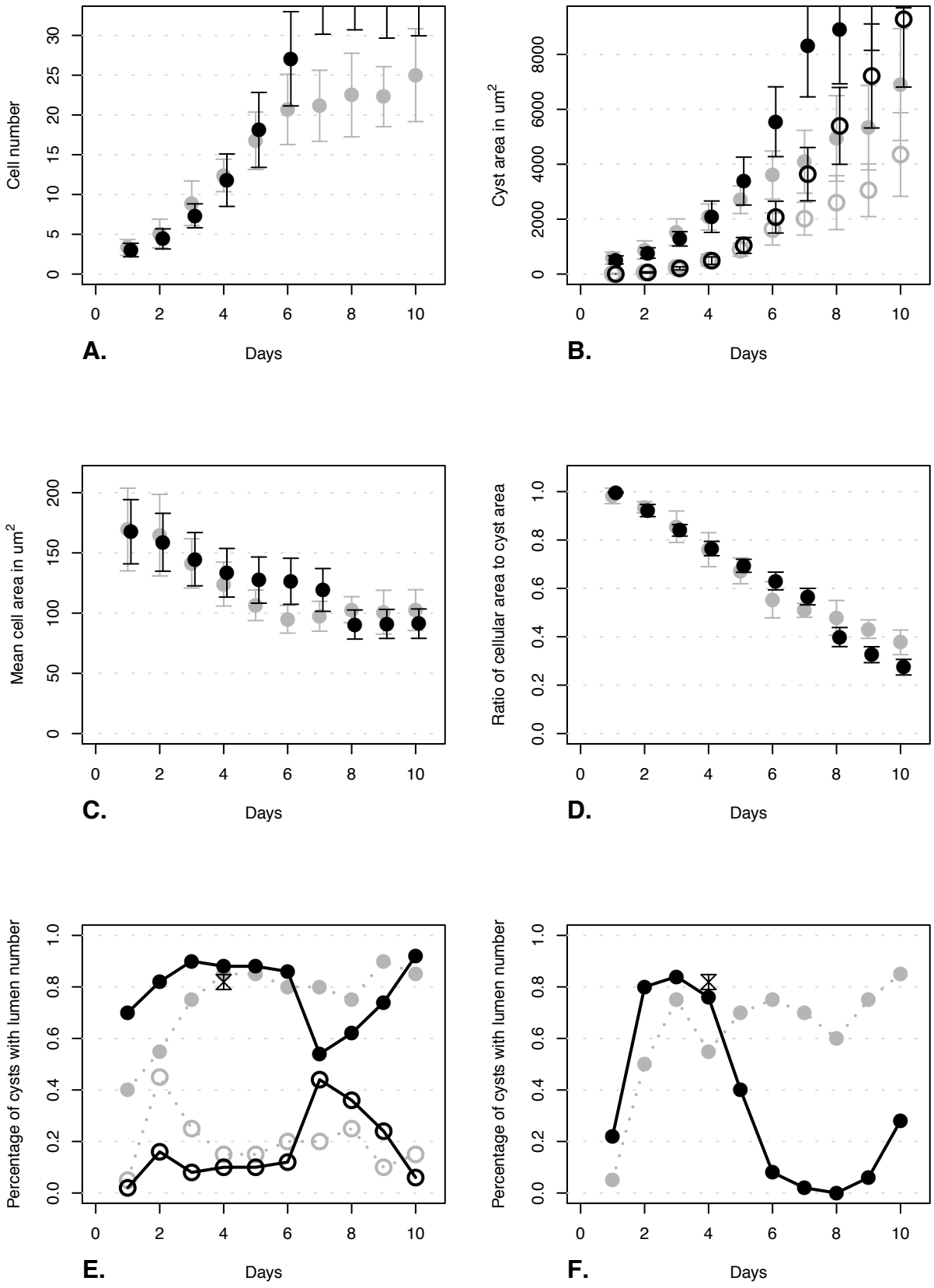
Appendix 1, Figure 37. Varied shiftDelay with stableRatio = 100000: sd = 100



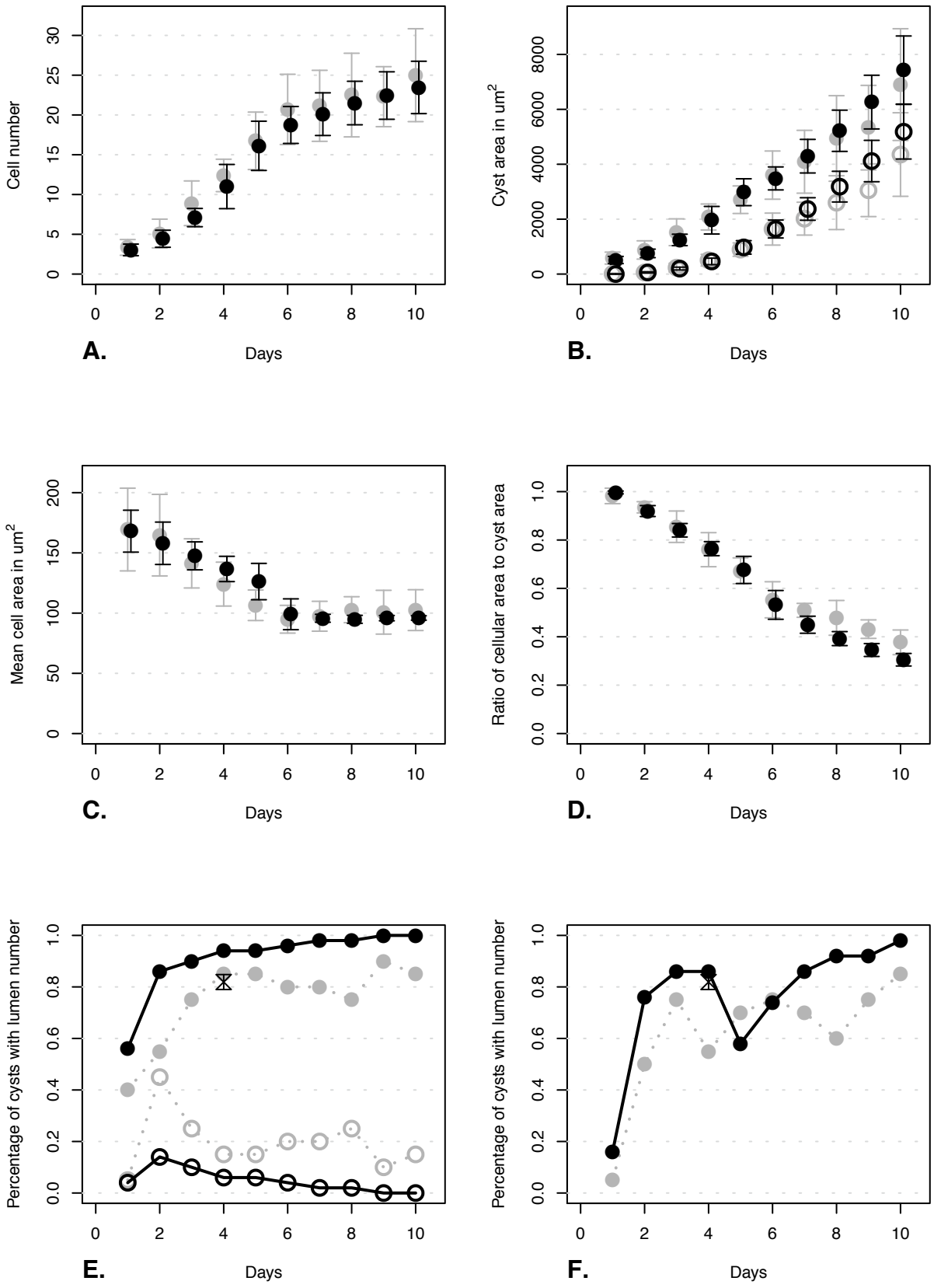
Appendix 1, Figure 38. Varied shiftDelay with stableRatio = 100000: sd = 250



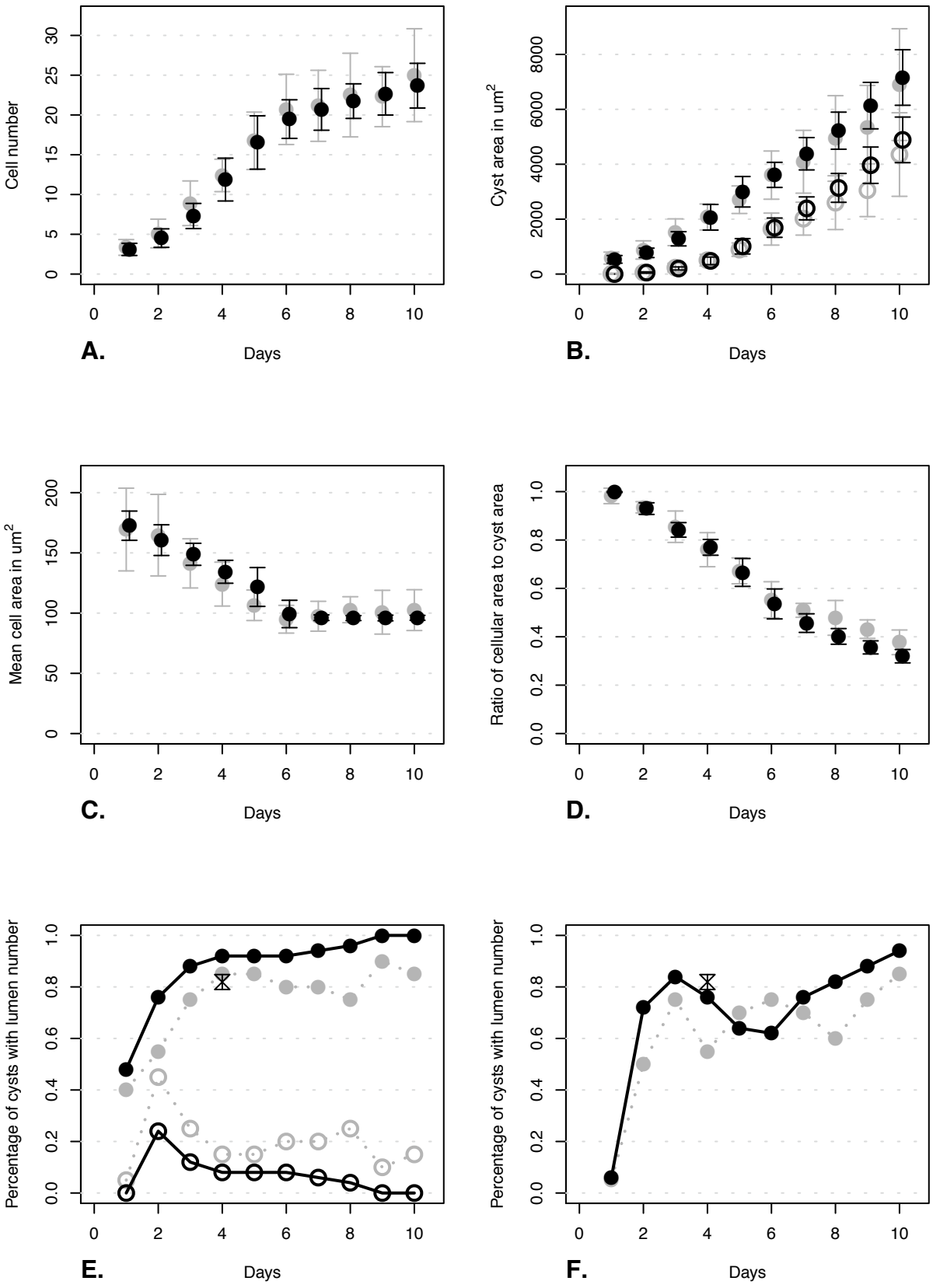
Appendix 1, Figure 39. Varied shiftDelay with stableRatio = 100000: sd = 300



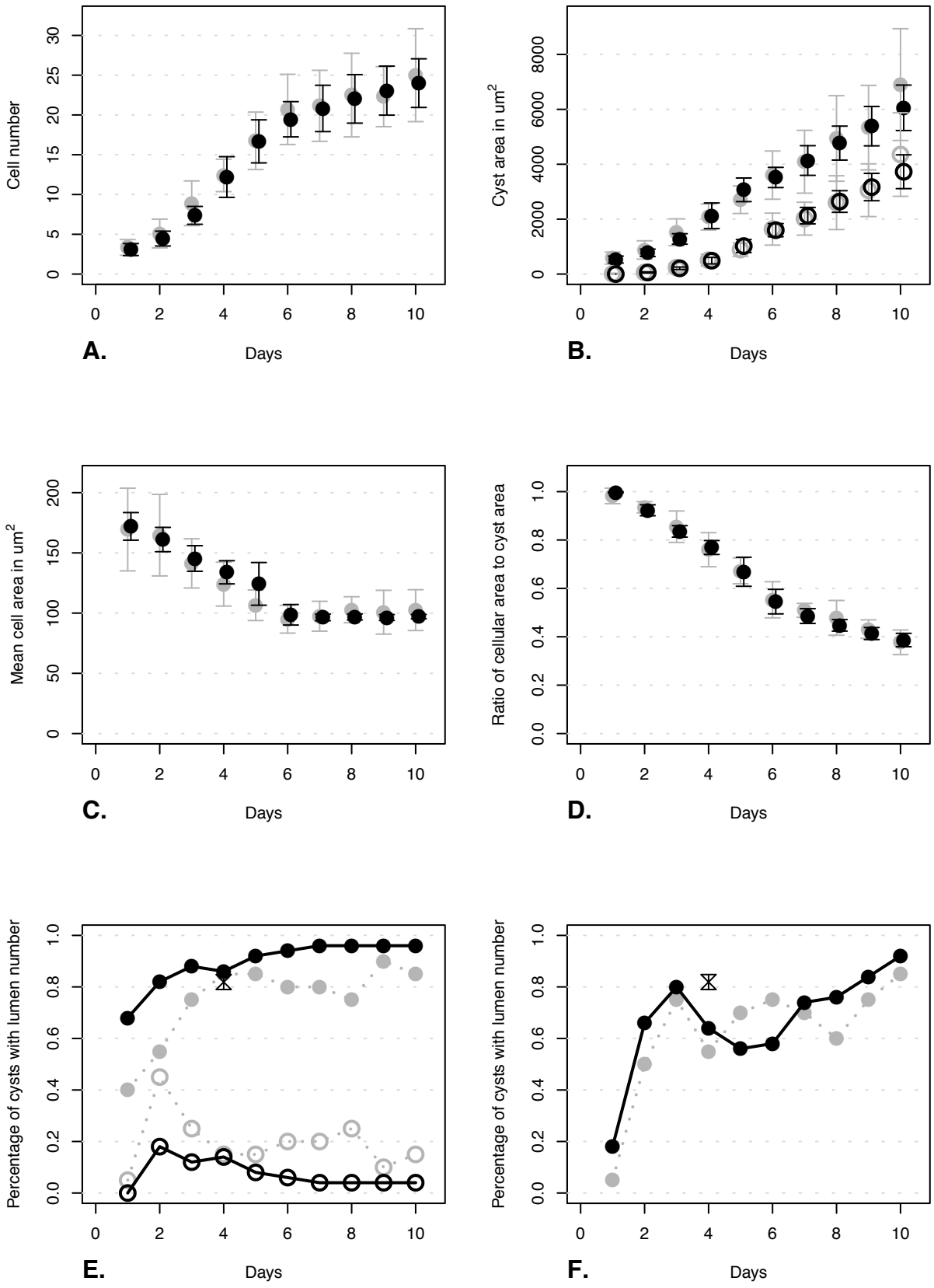
Appendix 1, Figure 40. Varied lgrSubtract: lgrs = 14



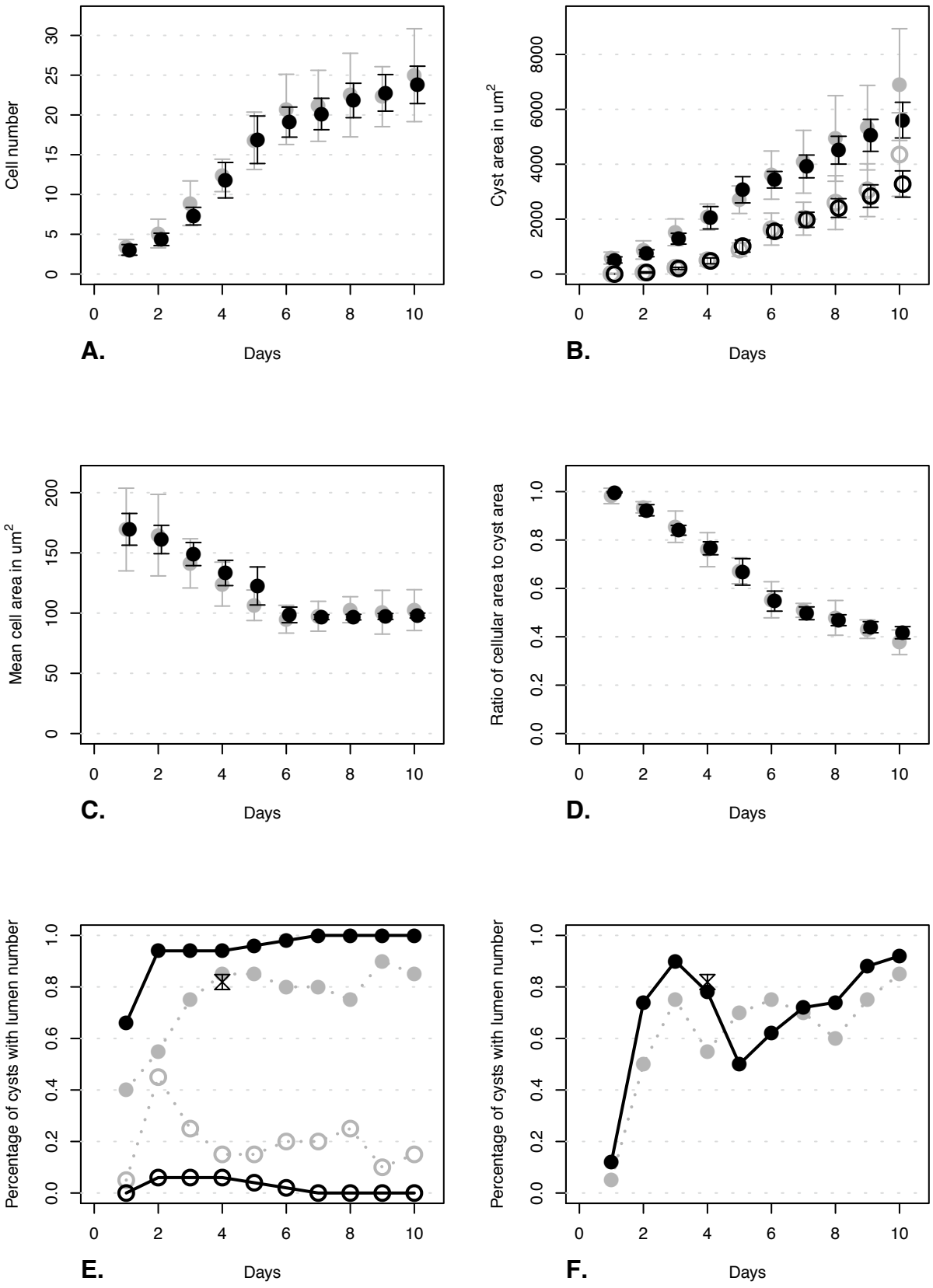
Appendix 1, Figure 41. Varied lgrSubtract: lgrs = 20



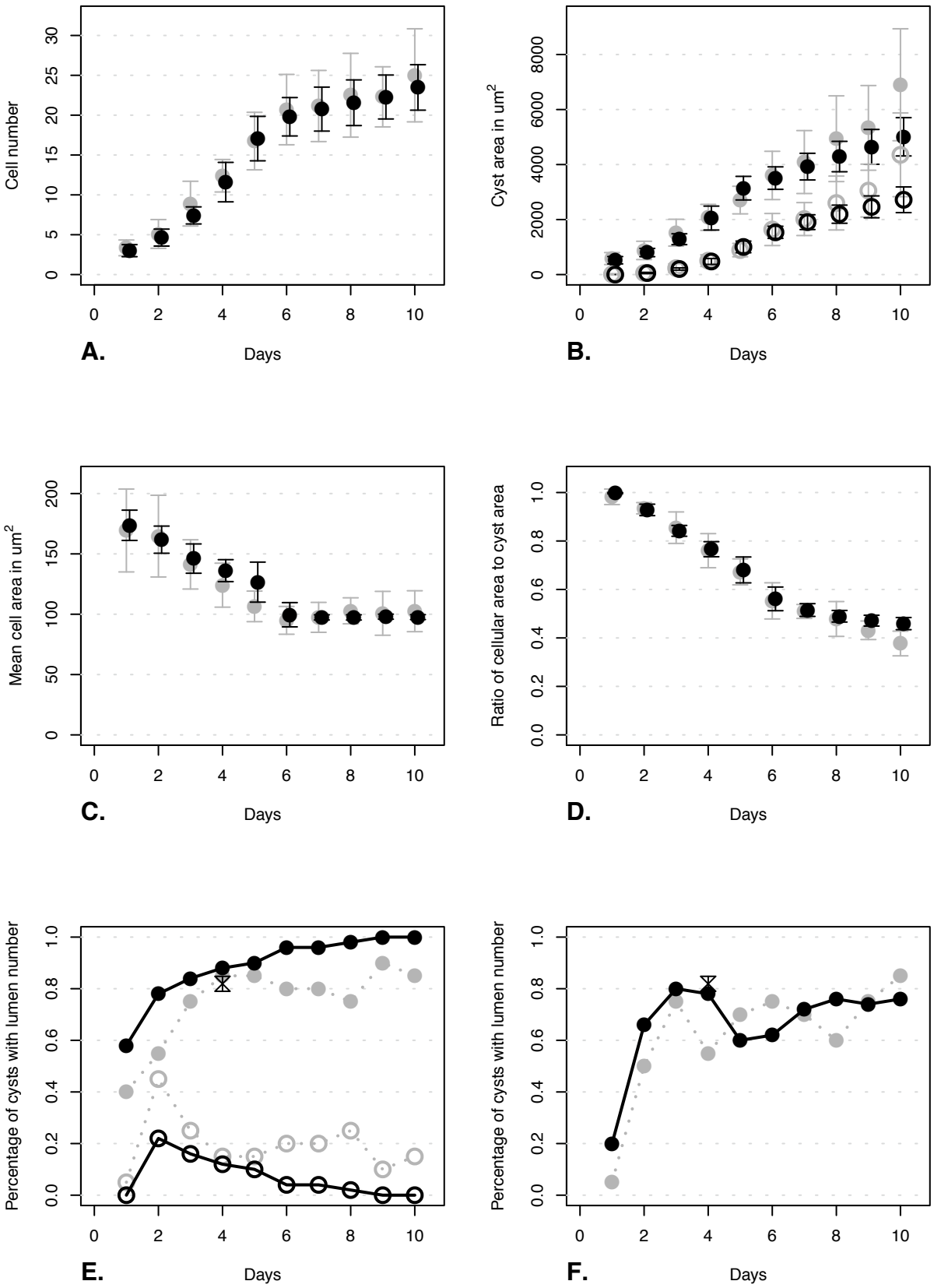
Appendix 1, Figure 42. Varied lgrSubtract: lgrs = 40



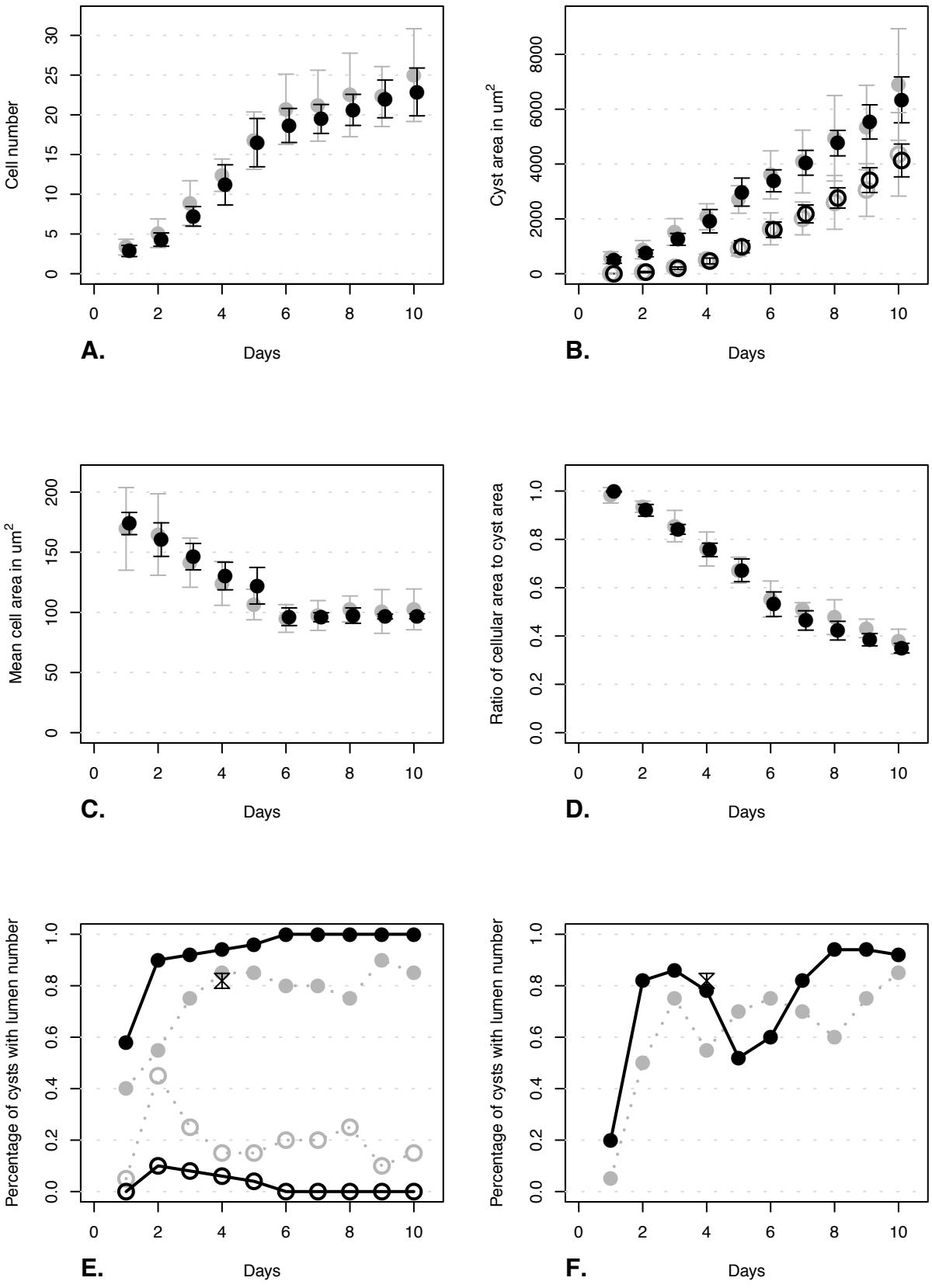
Appendix 1, Figure 43. Varied lgrSubtract: lgrs = 54



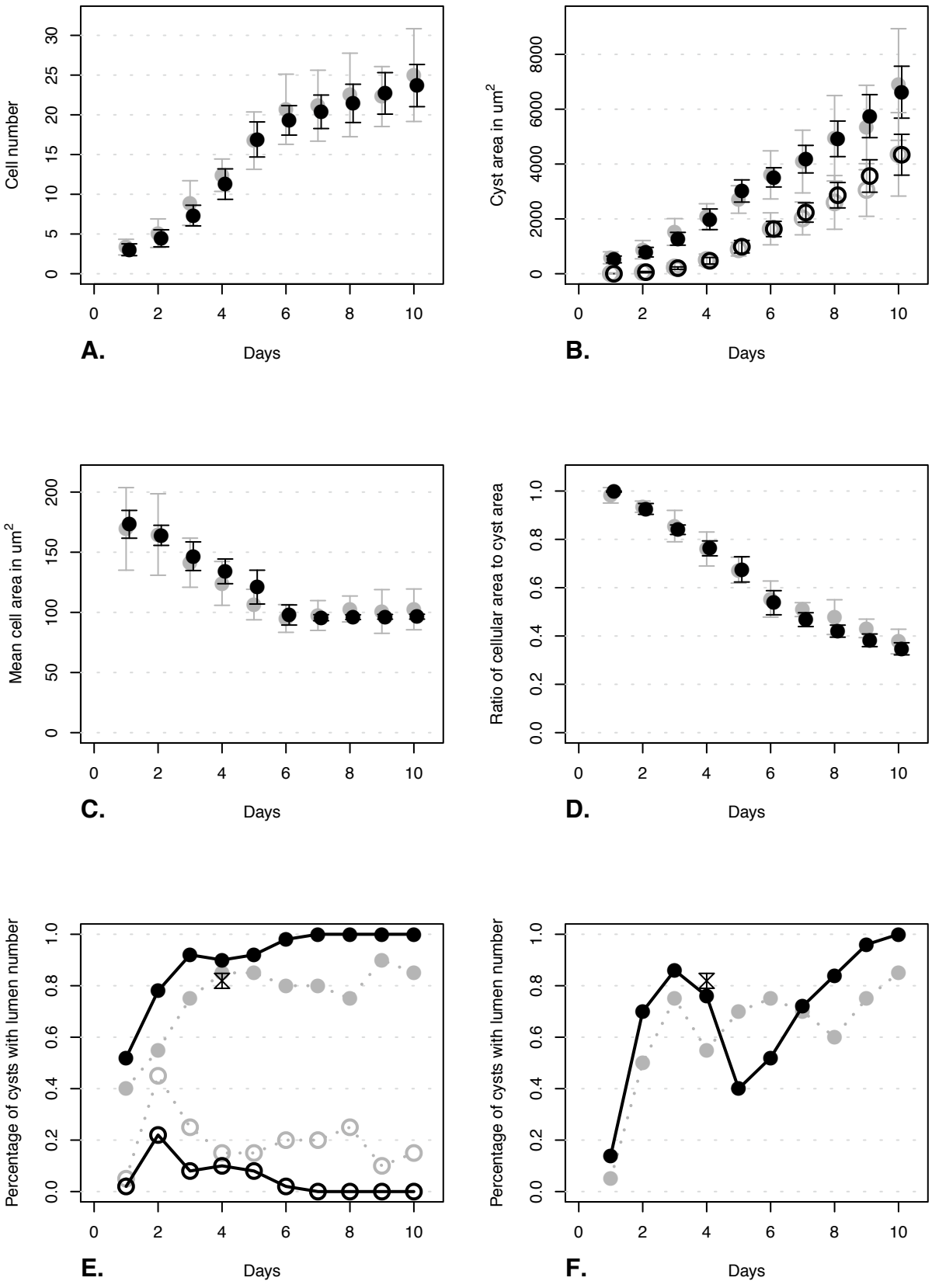
Appendix 1, Figure 44. Varied lgrSubtract: lgrs = 75



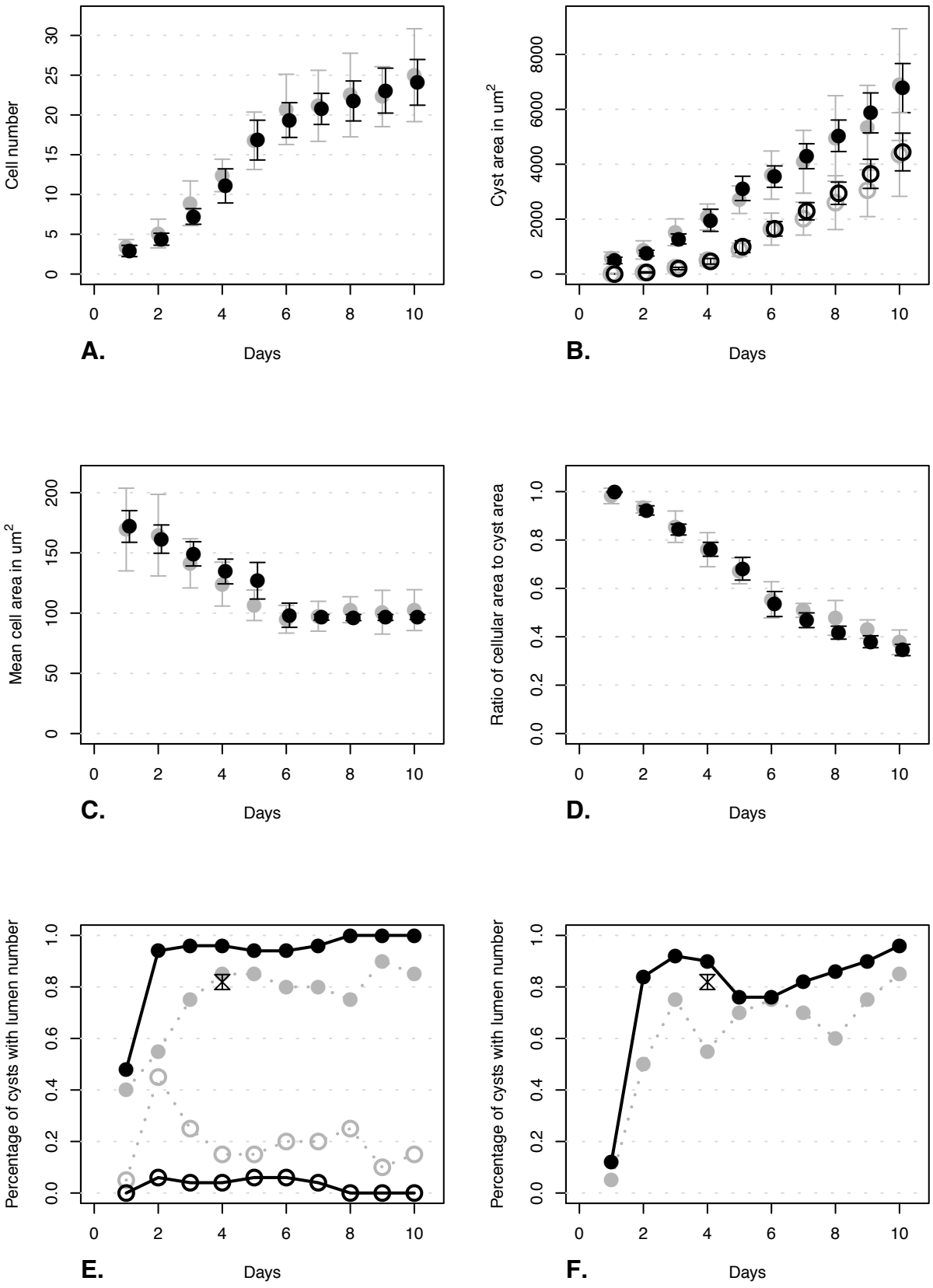
Appendix 1, Figure 45. Varied doublingArea: da = 20



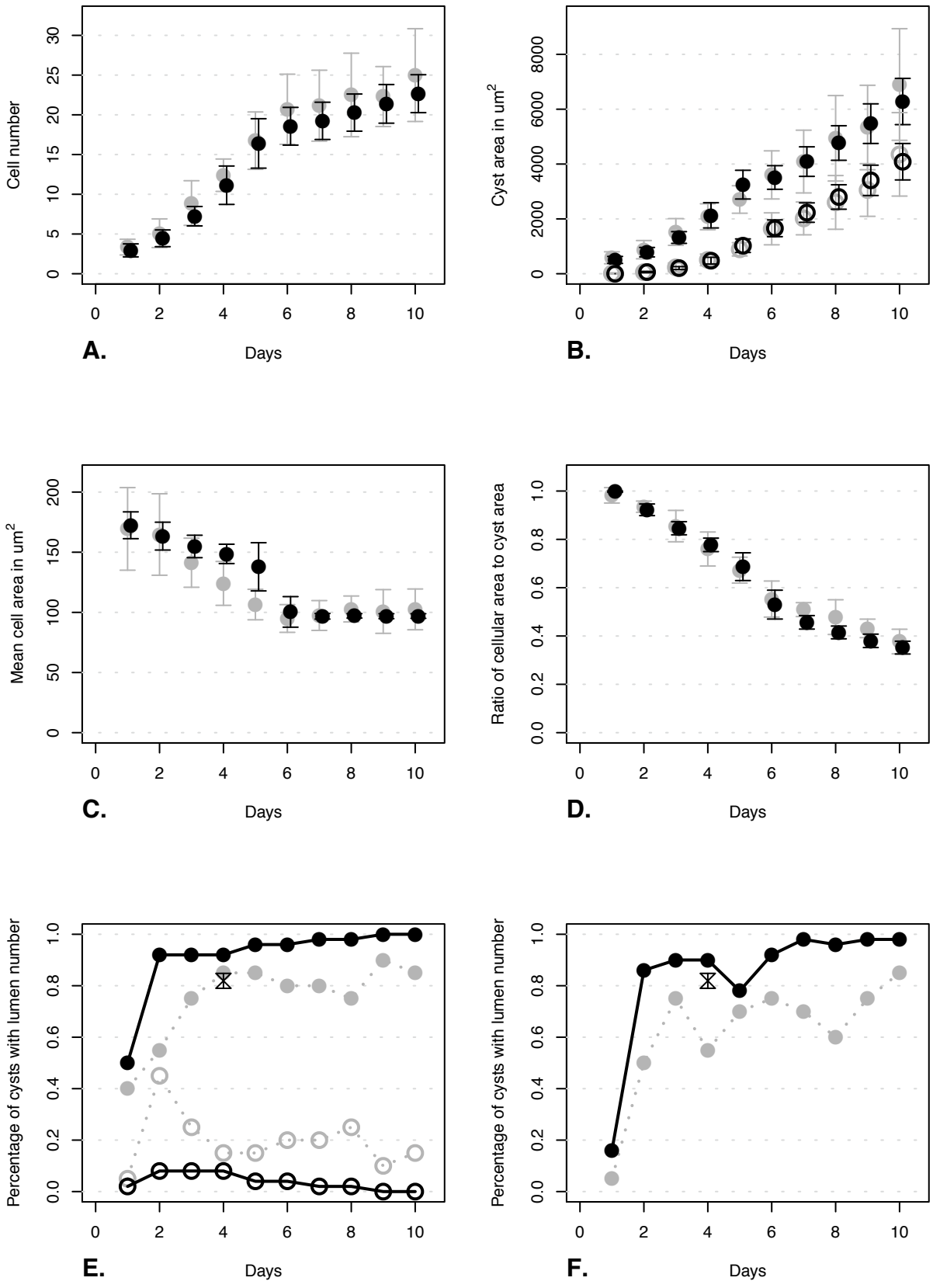
Appendix 1, Figure 46. Varied doublingArea: da = 30



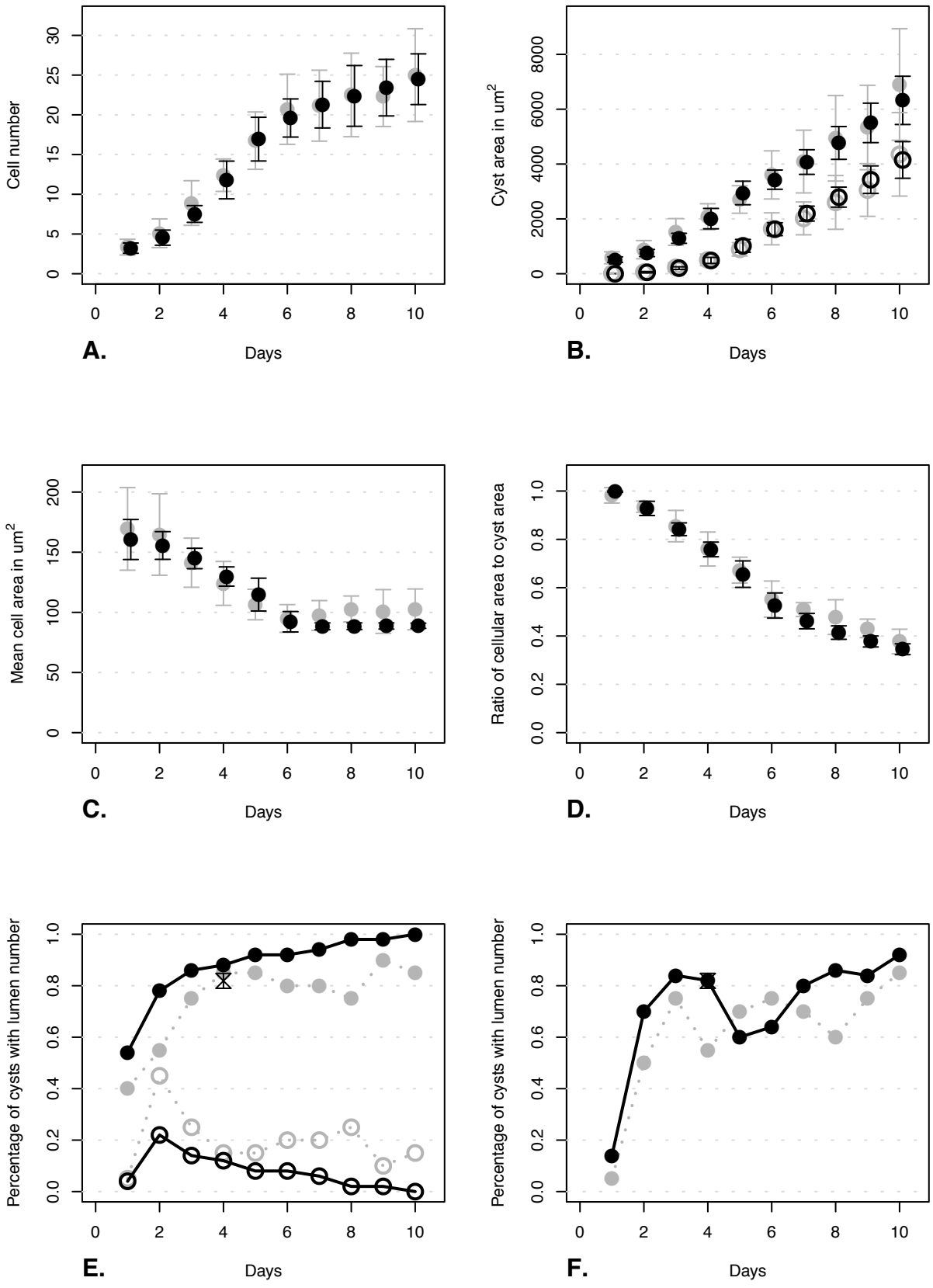
Appendix 1, Figure 47. Varied doublingArea: da = 50



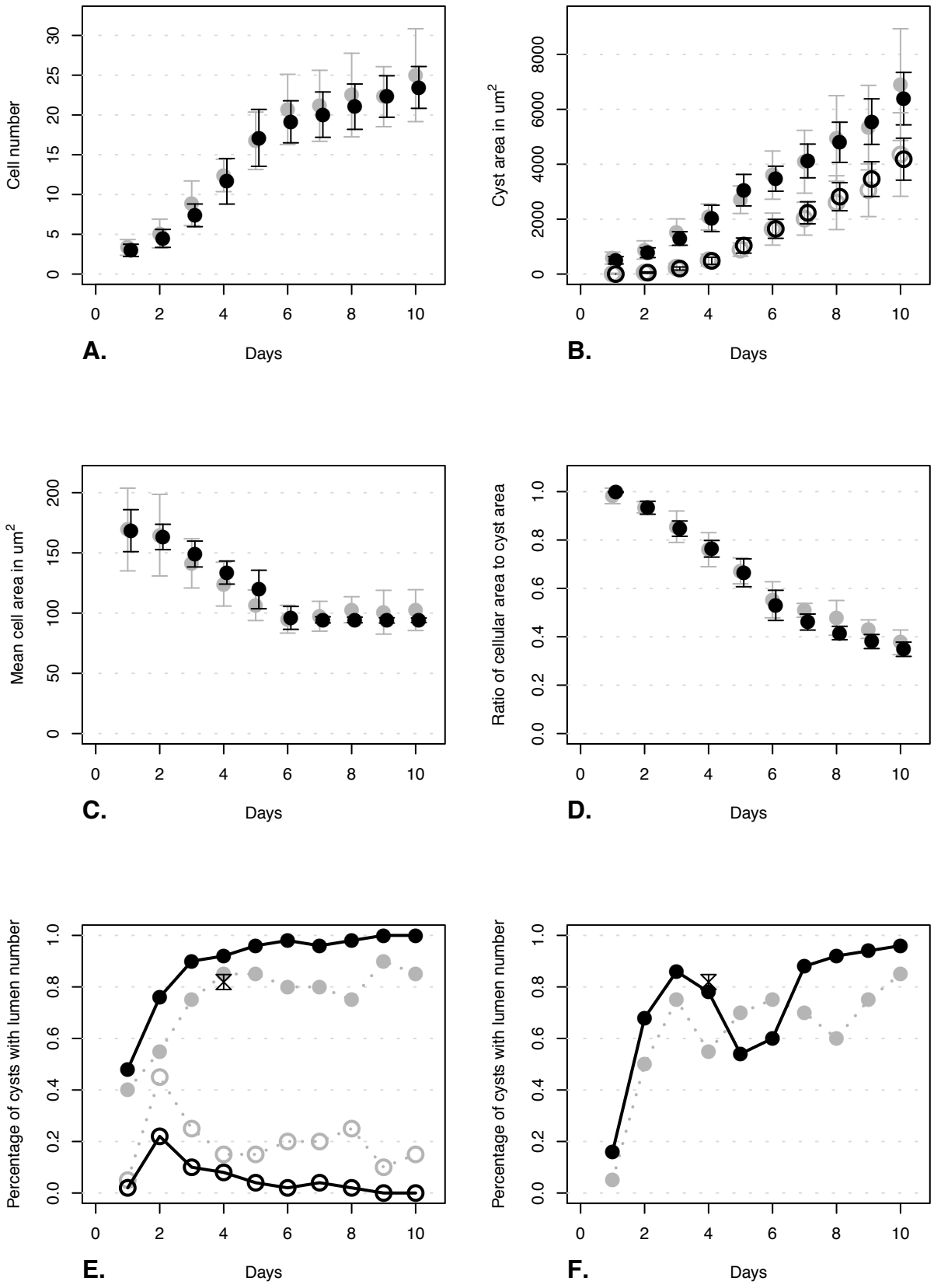
Appendix 1, Figure 48. Varied doublingArea: da = 80



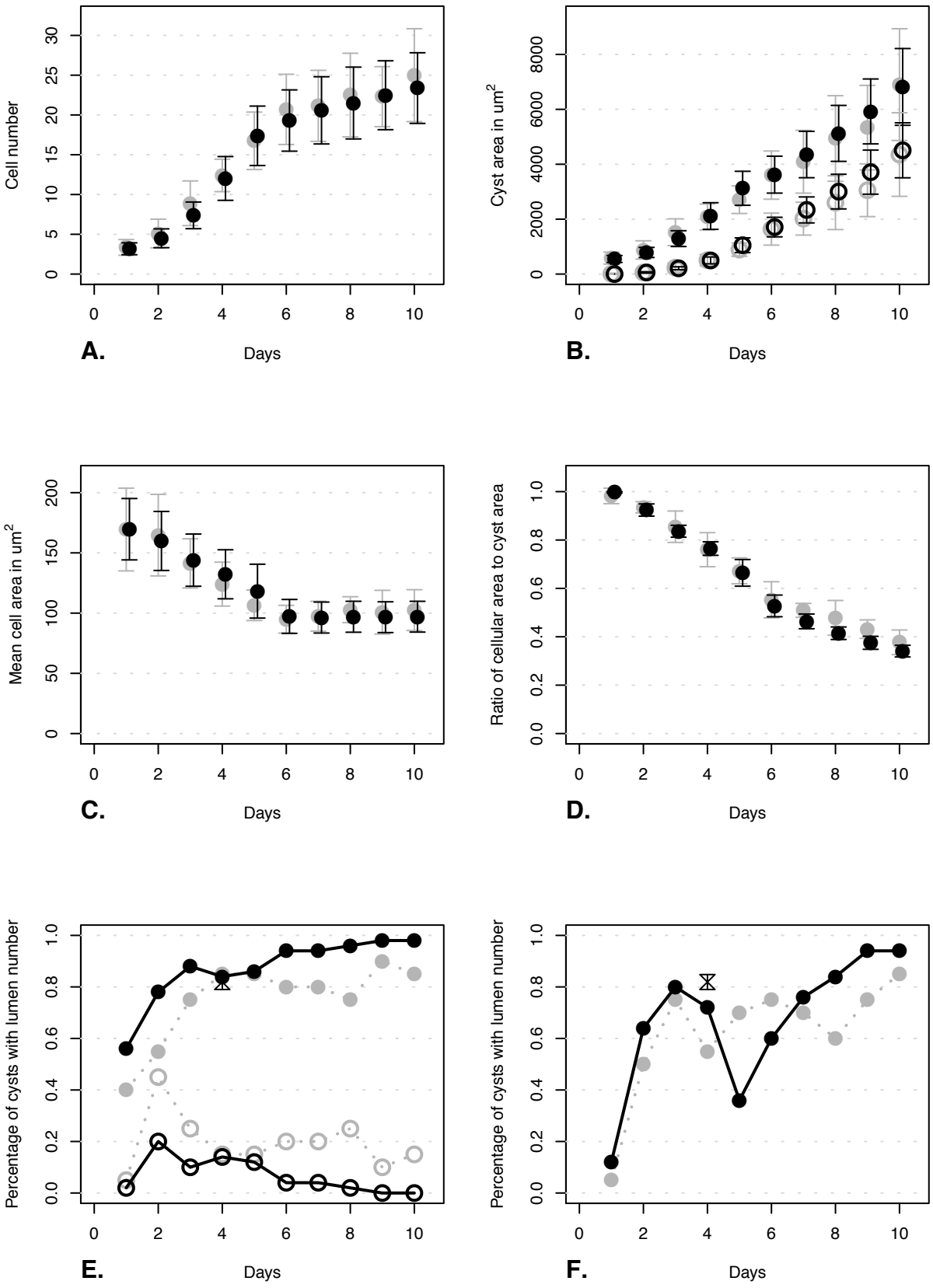
Appendix 1, Figure 49. Varied multiplier: m = 0.3



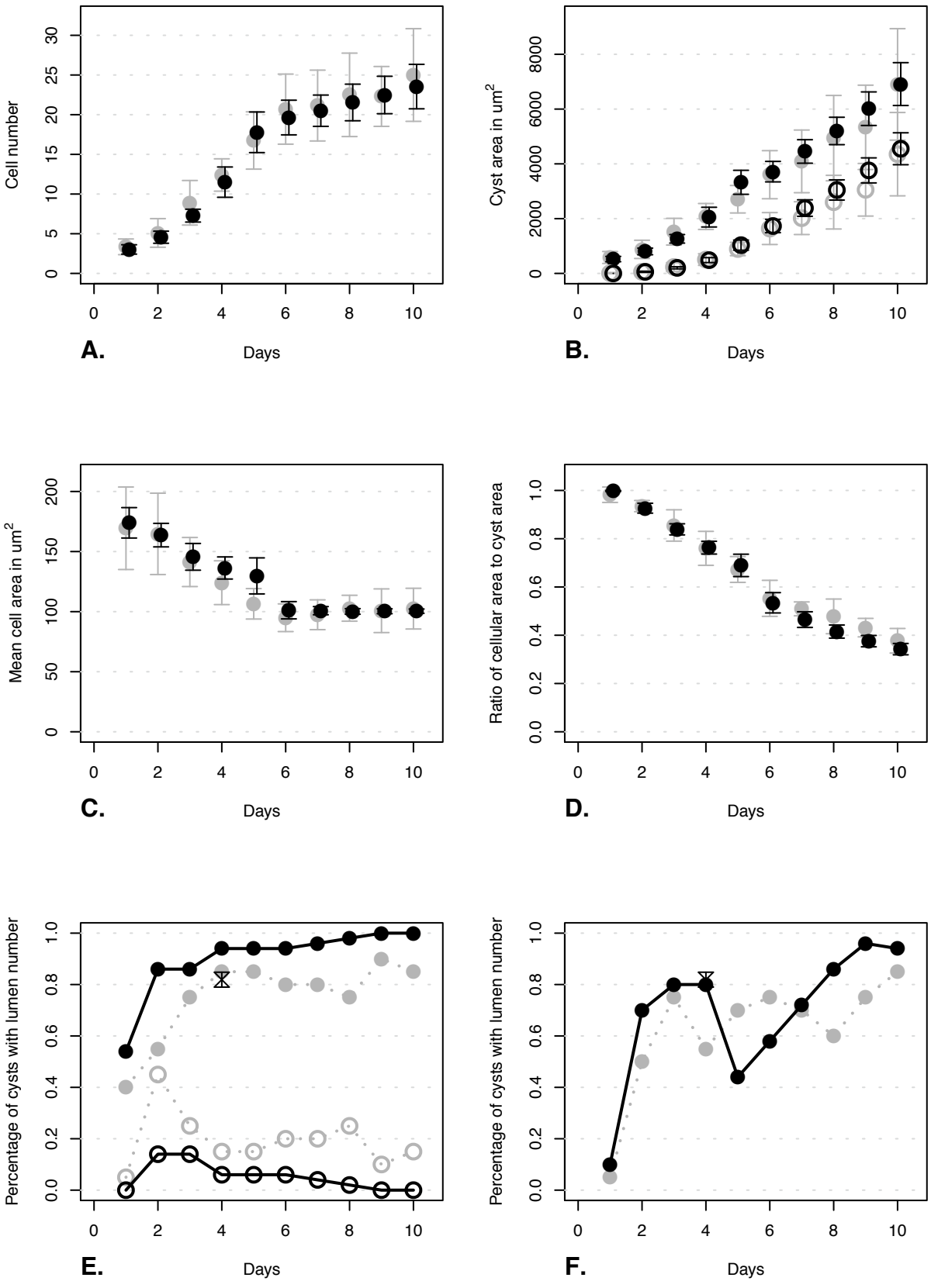
Appendix 1, Figure 50. Varied multiplier: m = 0.5



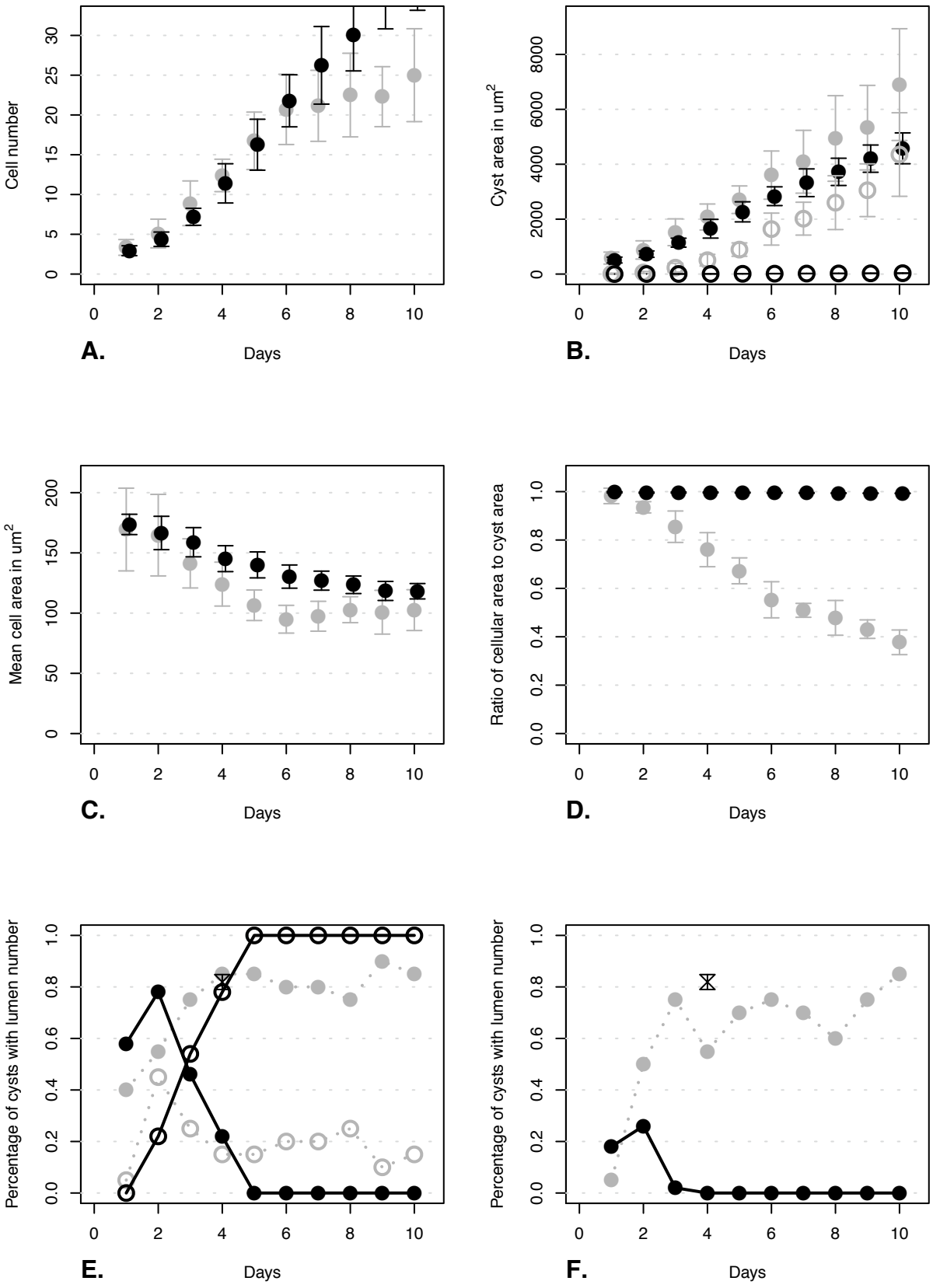
Appendix 1, Figure 51. Varied multiplier: m = 0.7



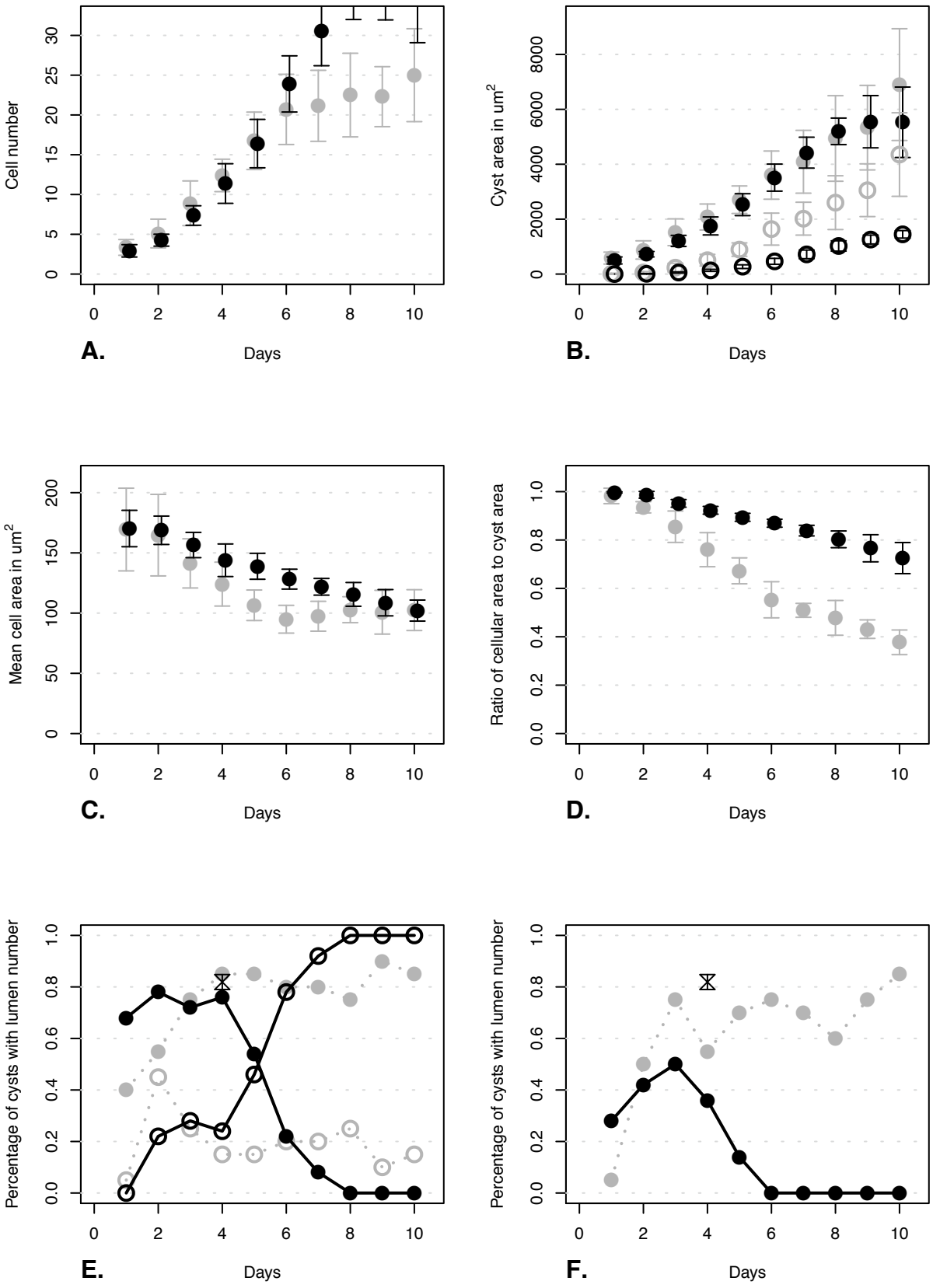
Appendix 1, Figure 52. Varied multiplier: m = 0.8



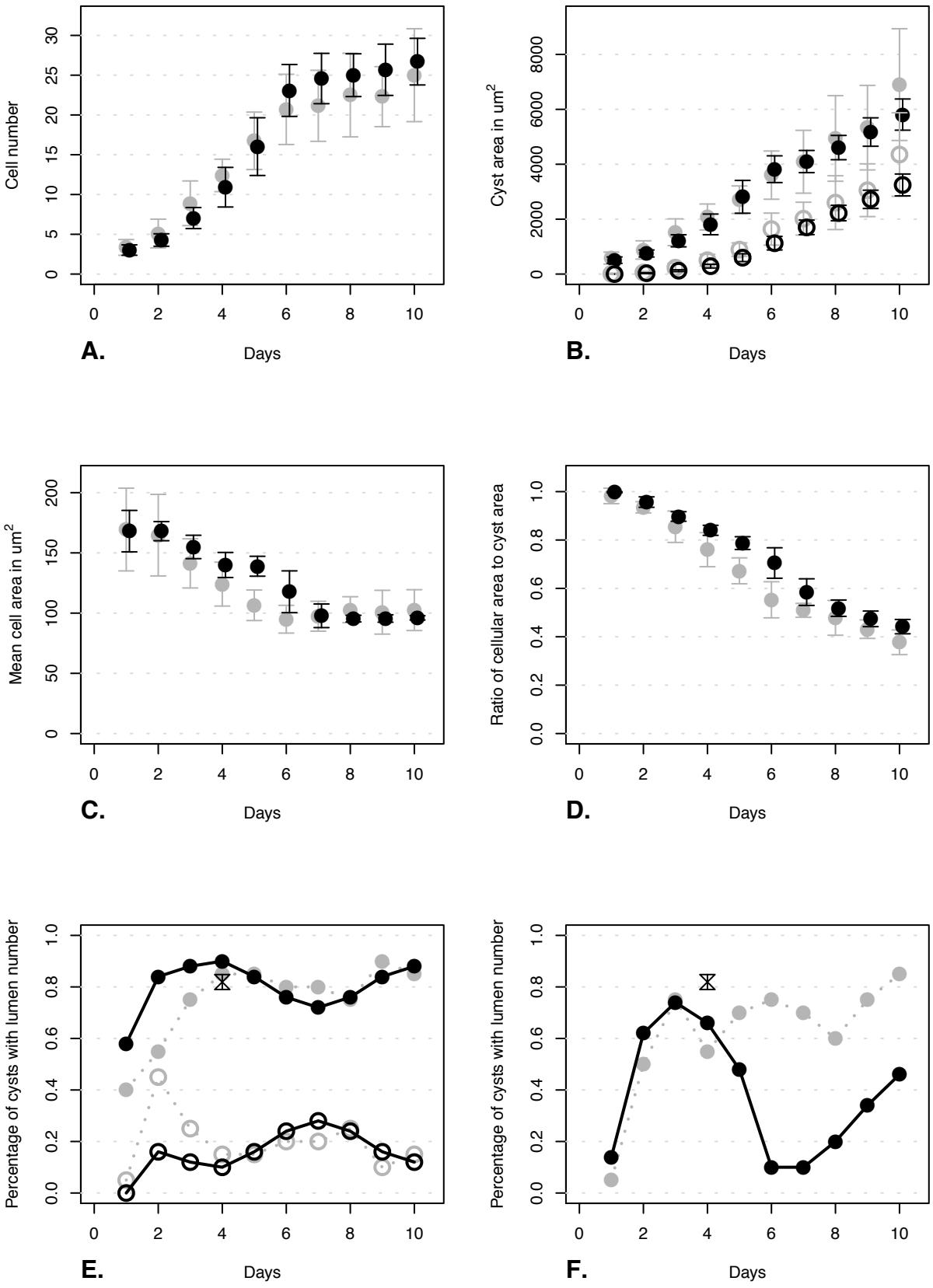
Appendix 1, Figure 53. Varied lumenGrowthRate: lgr = 0



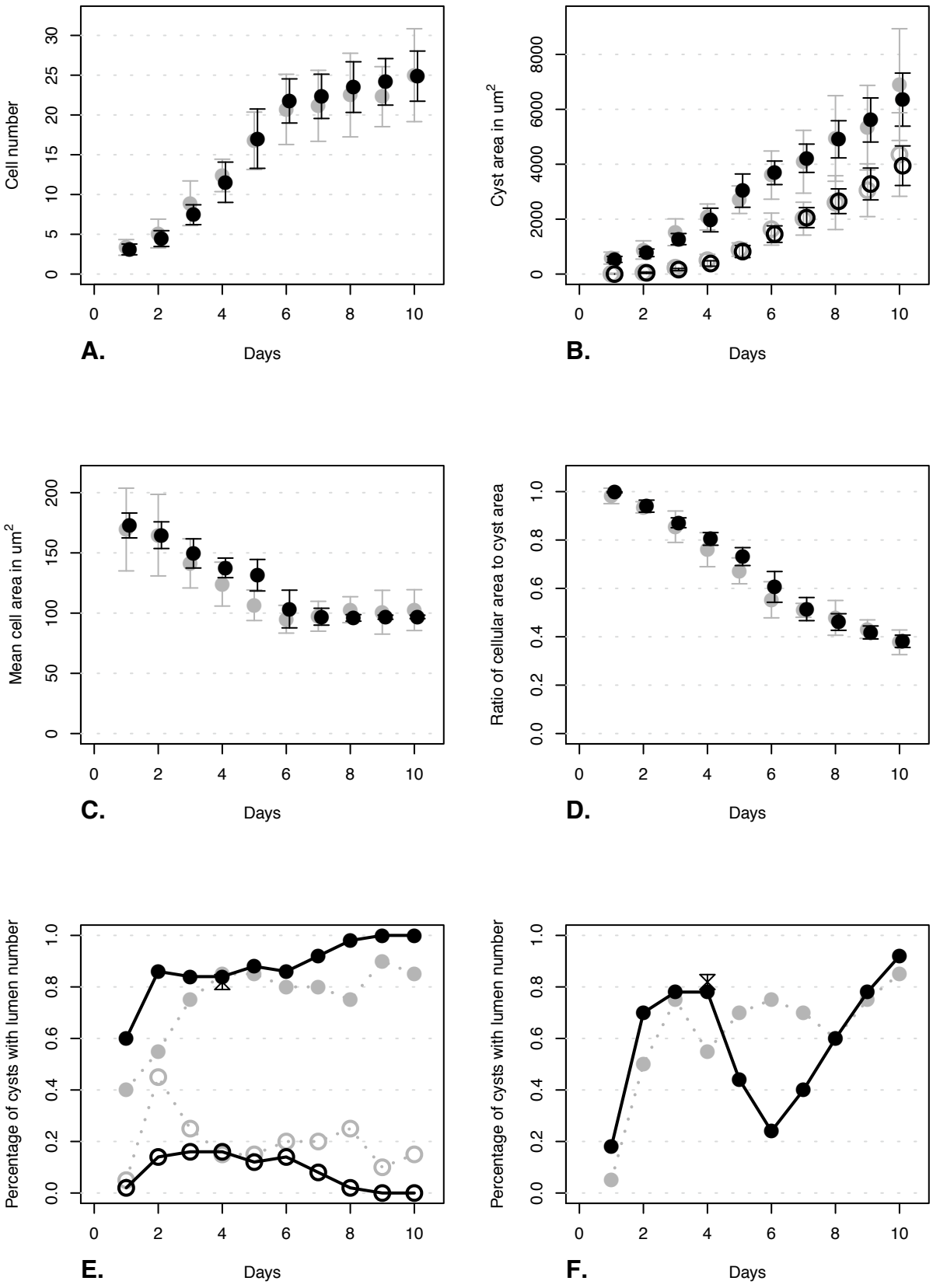
Appendix 1, Figure 54. Varied lumenGrowthRate: lgr = 0.001



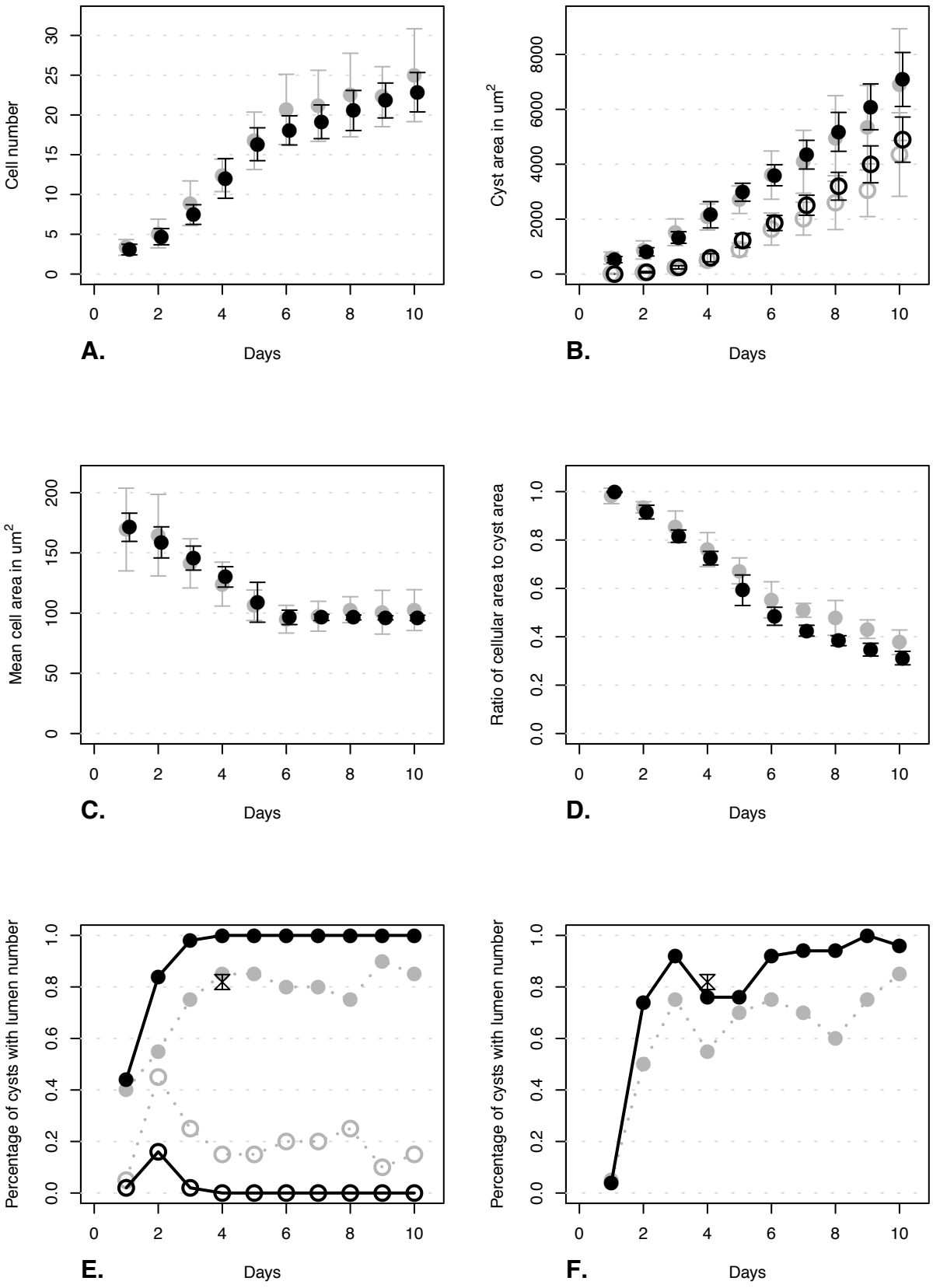
Appendix 1, Figure 55. Varied lumenGrowthRate: lgr = 0.002



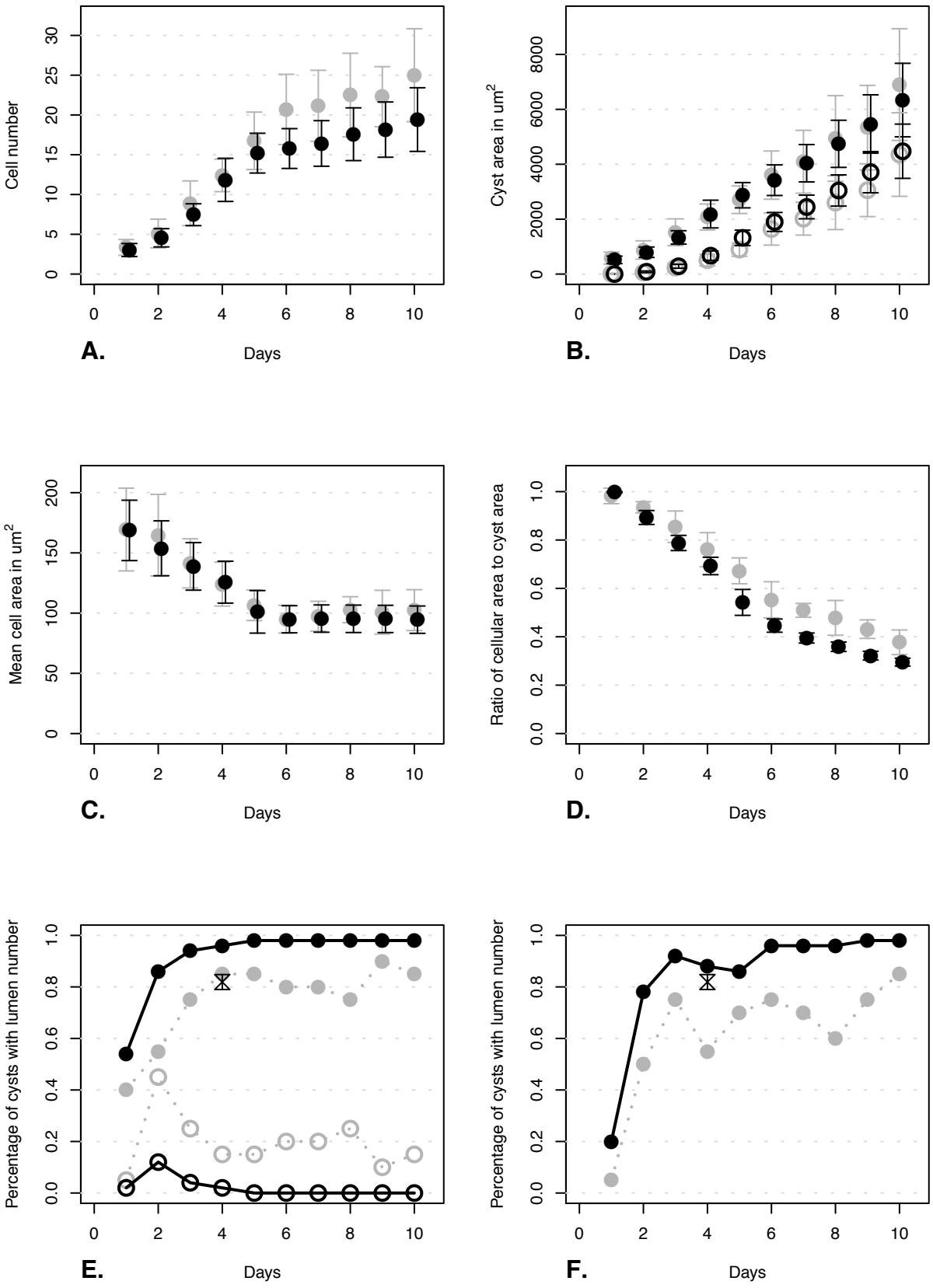
Appendix 1, Figure 56. Varied lumenGrowthRate: lgr = 0.0025



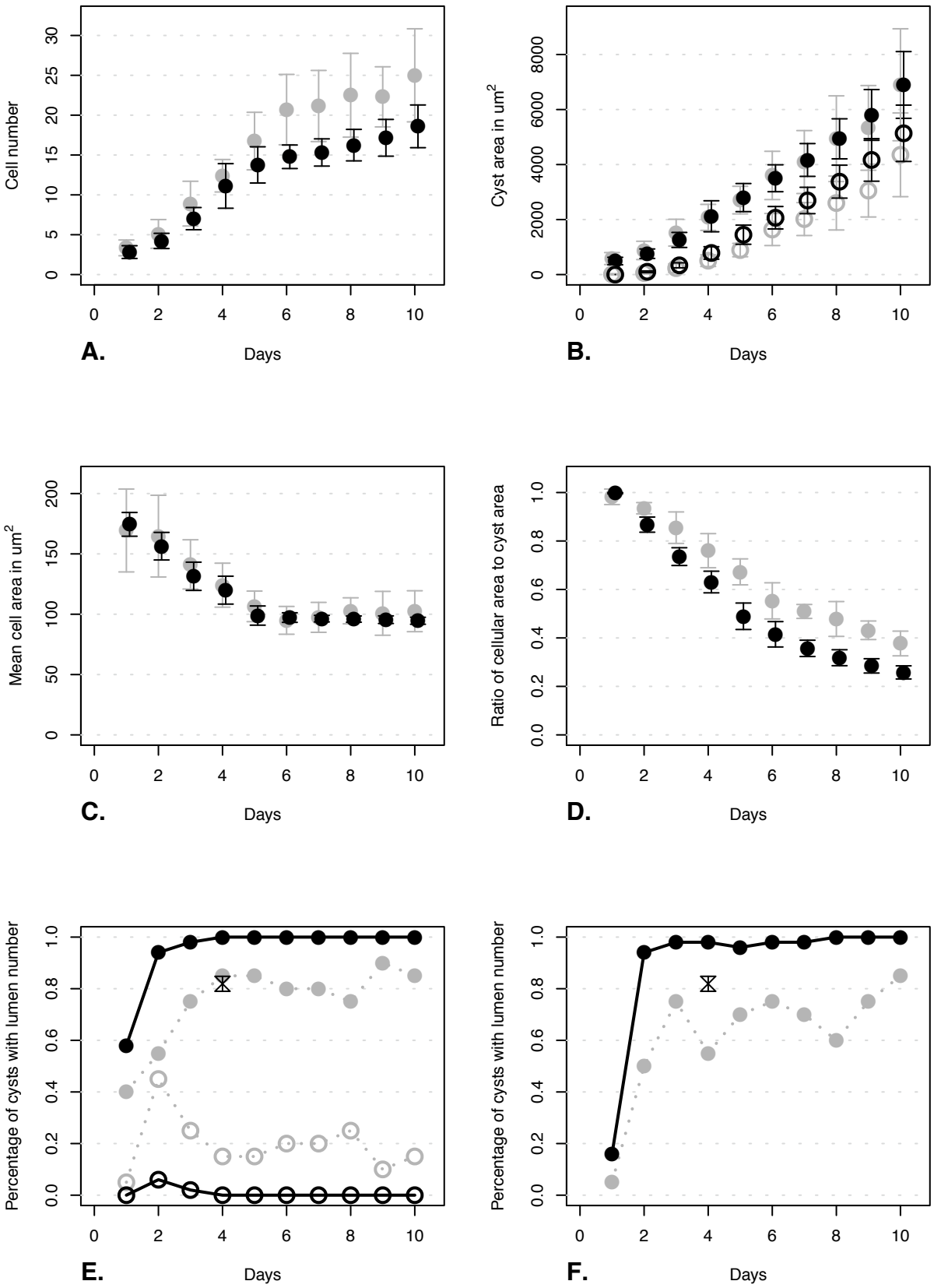
Appendix 1, Figure 57. Varied lumenGrowthRate: lgr = 0.0035



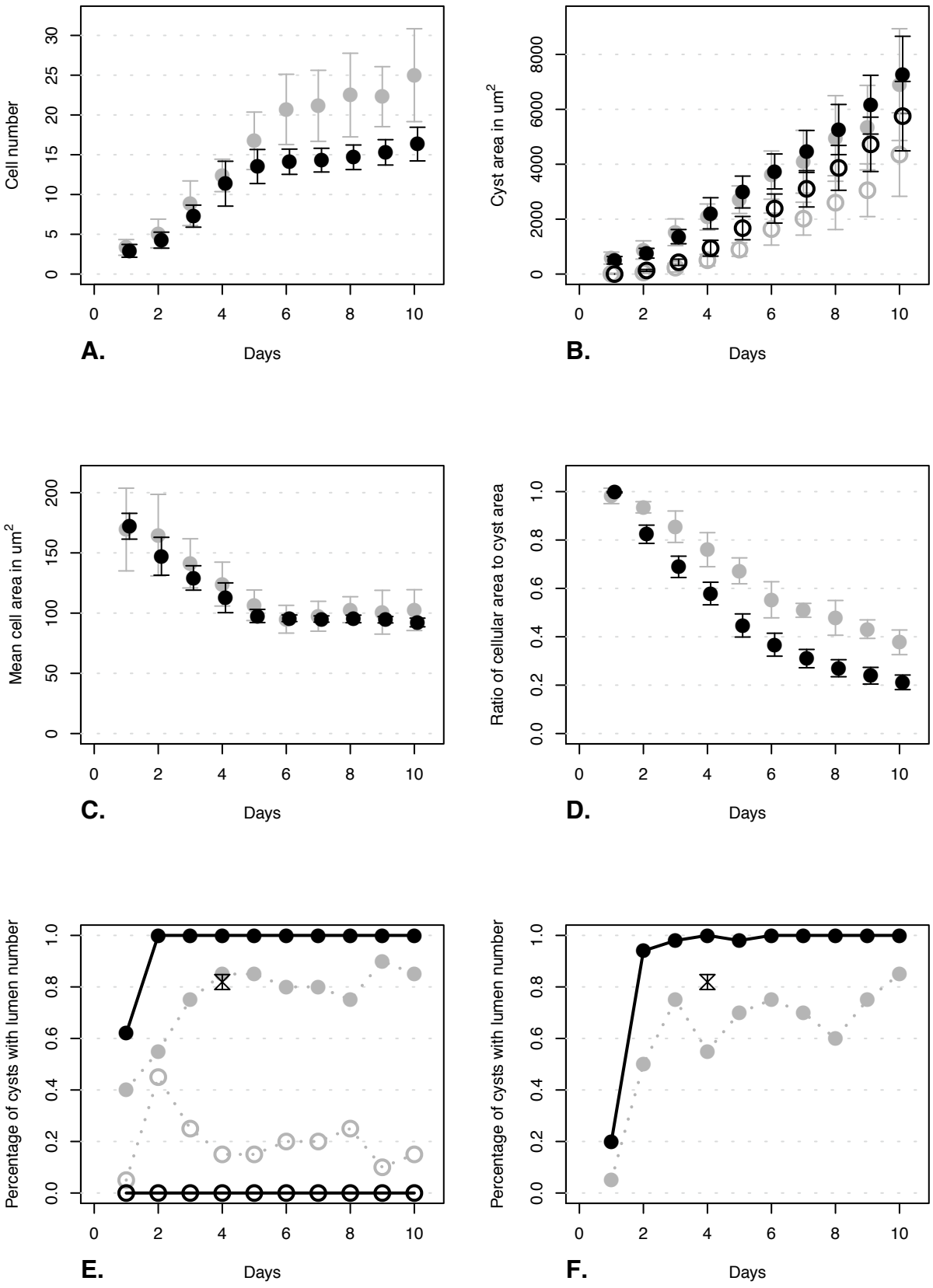
Appendix 1, Figure 58. Varied lumenGrowthRate: lgr = 0.004



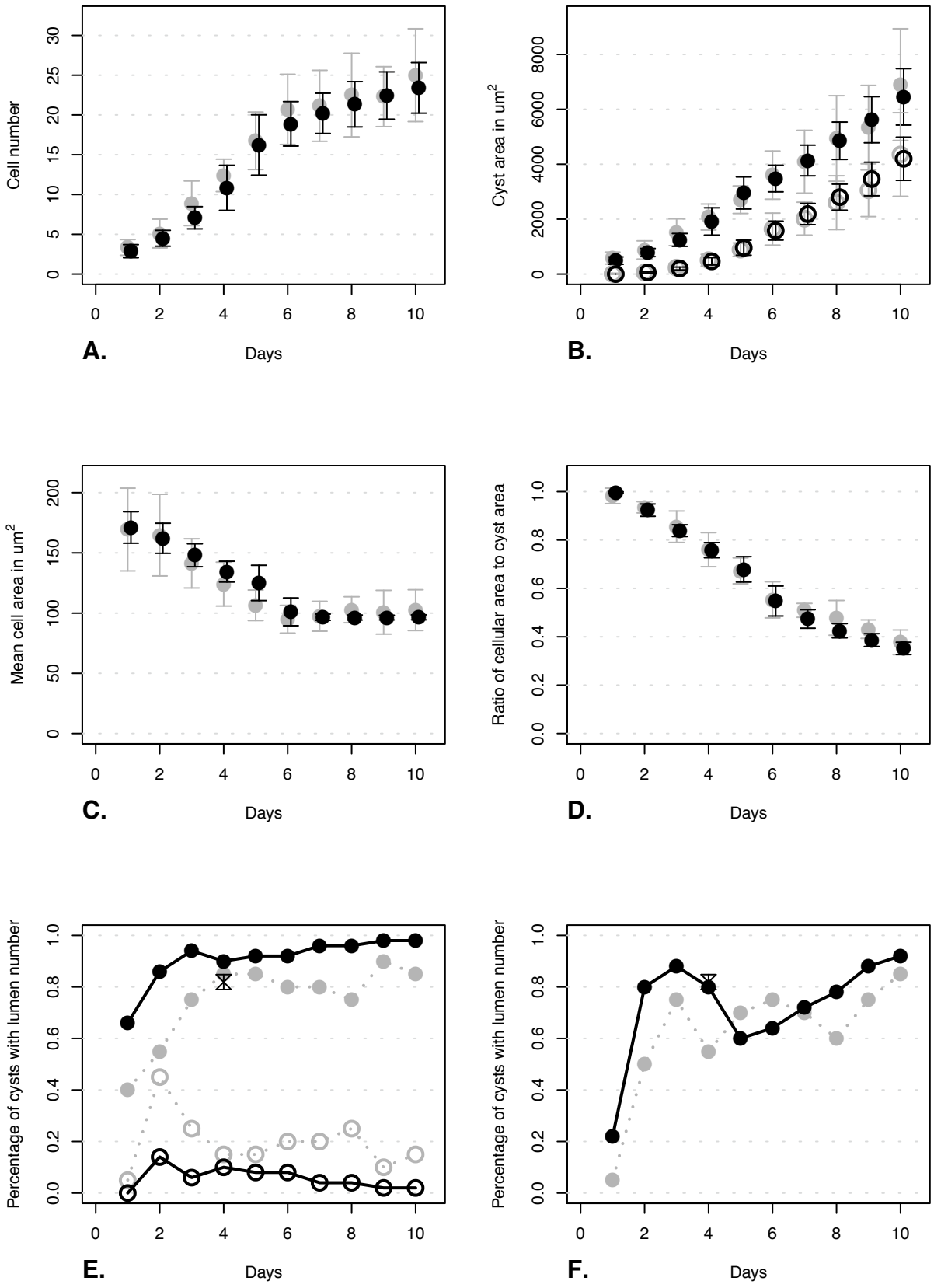
Appendix 1, Figure 59. Varied lumenGrowthRate: lgr = 0.005



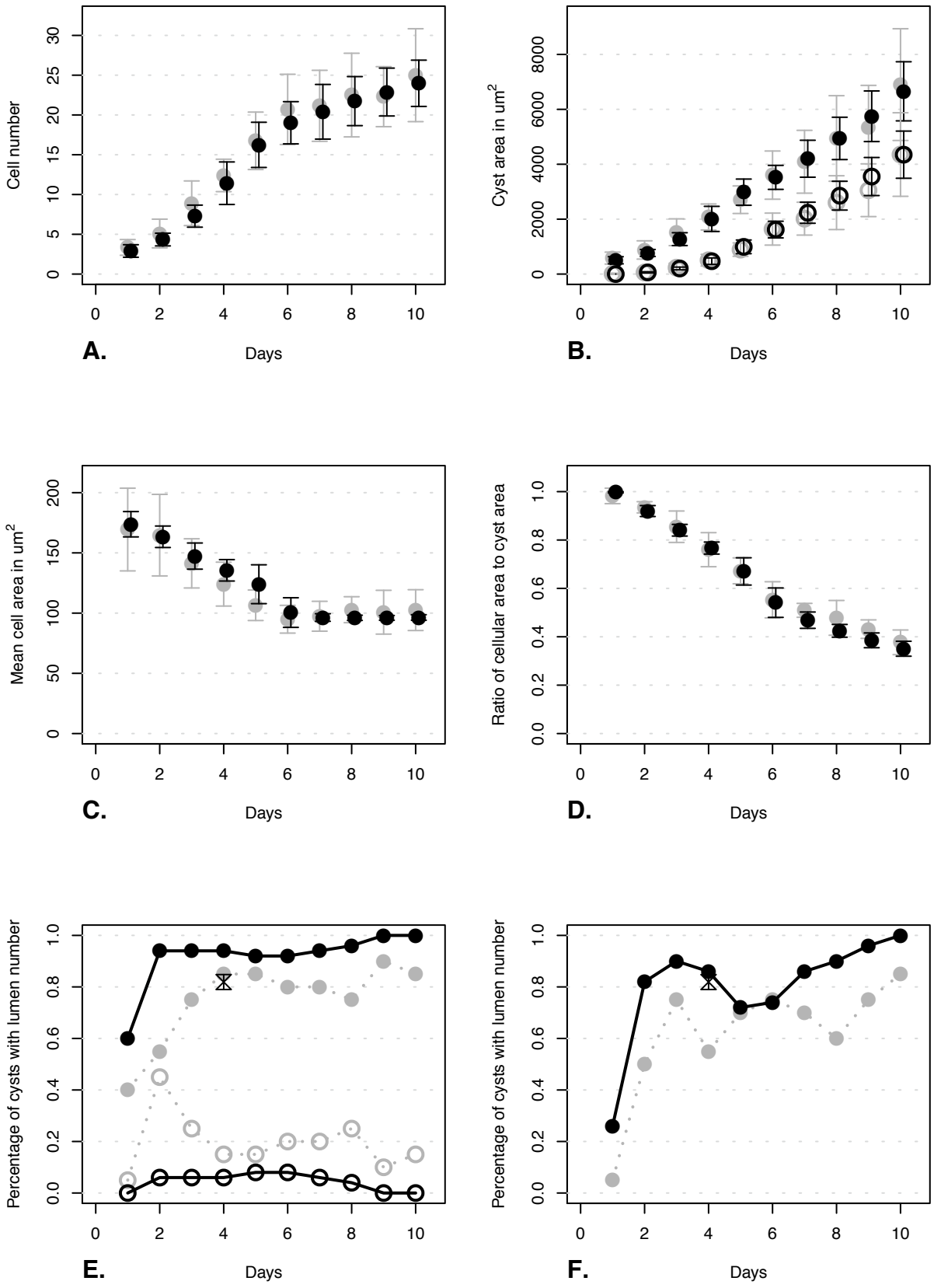
Appendix 1, Figure 60. Varied lumenGrowthRate: lgr = 0.006



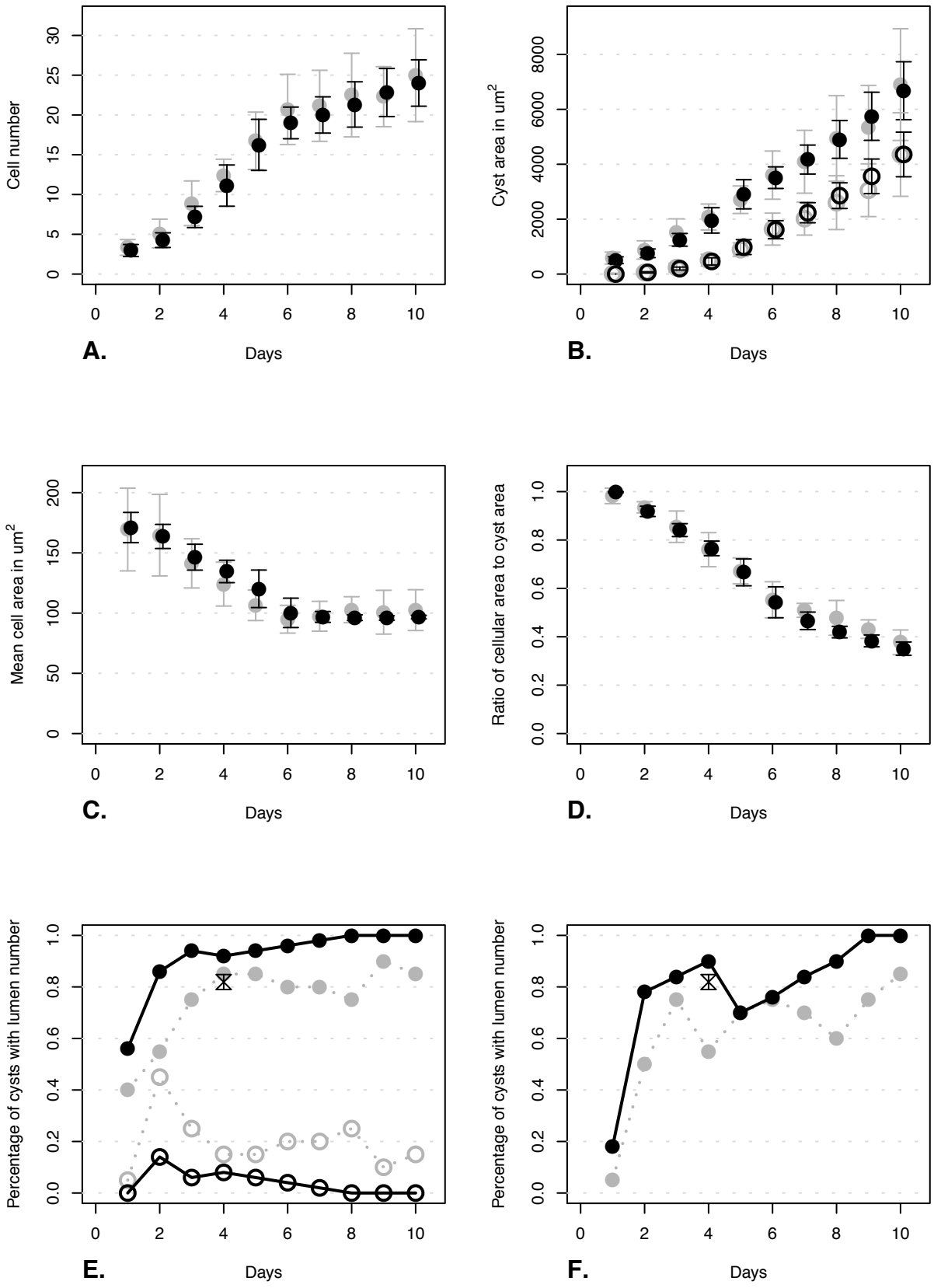
Appendix 1, Figure 61. Varied deathRateLumen: drl = 0.01



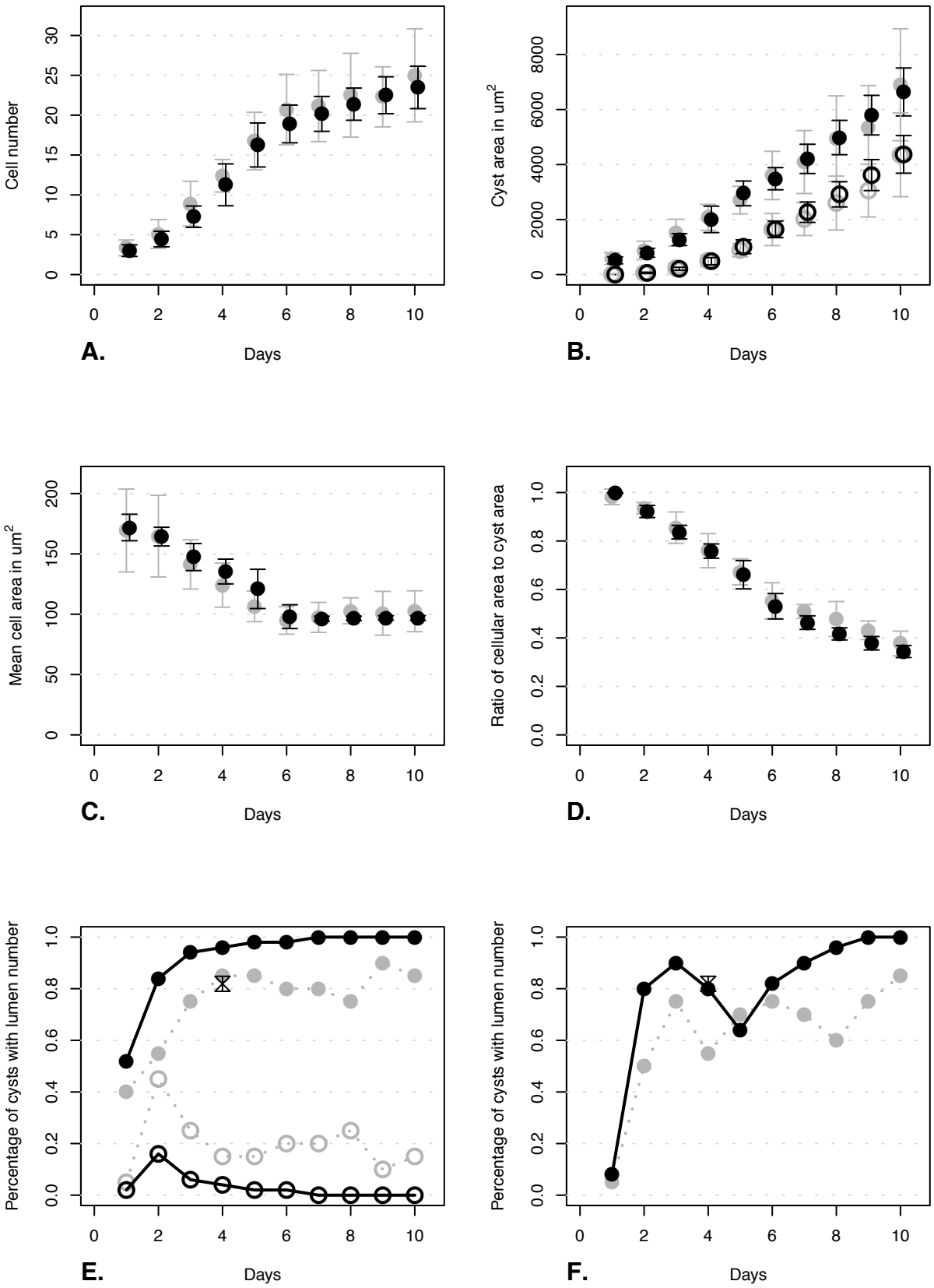
Appendix 1, Figure 62. Varied deathRateLumen: drl = 0.03



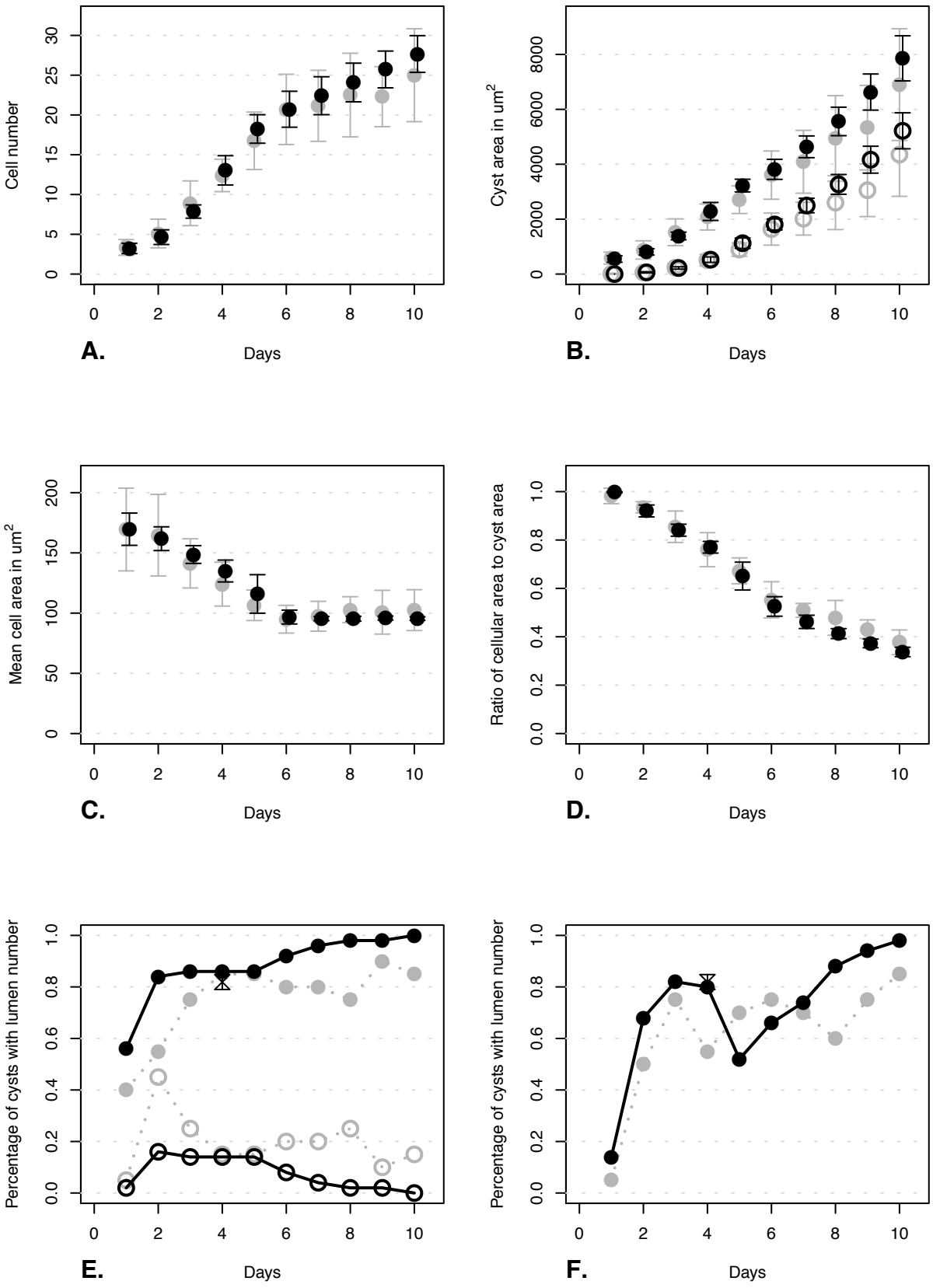
Appendix 1, Figure 63. Varied deathRateLumen: drl = 0.04



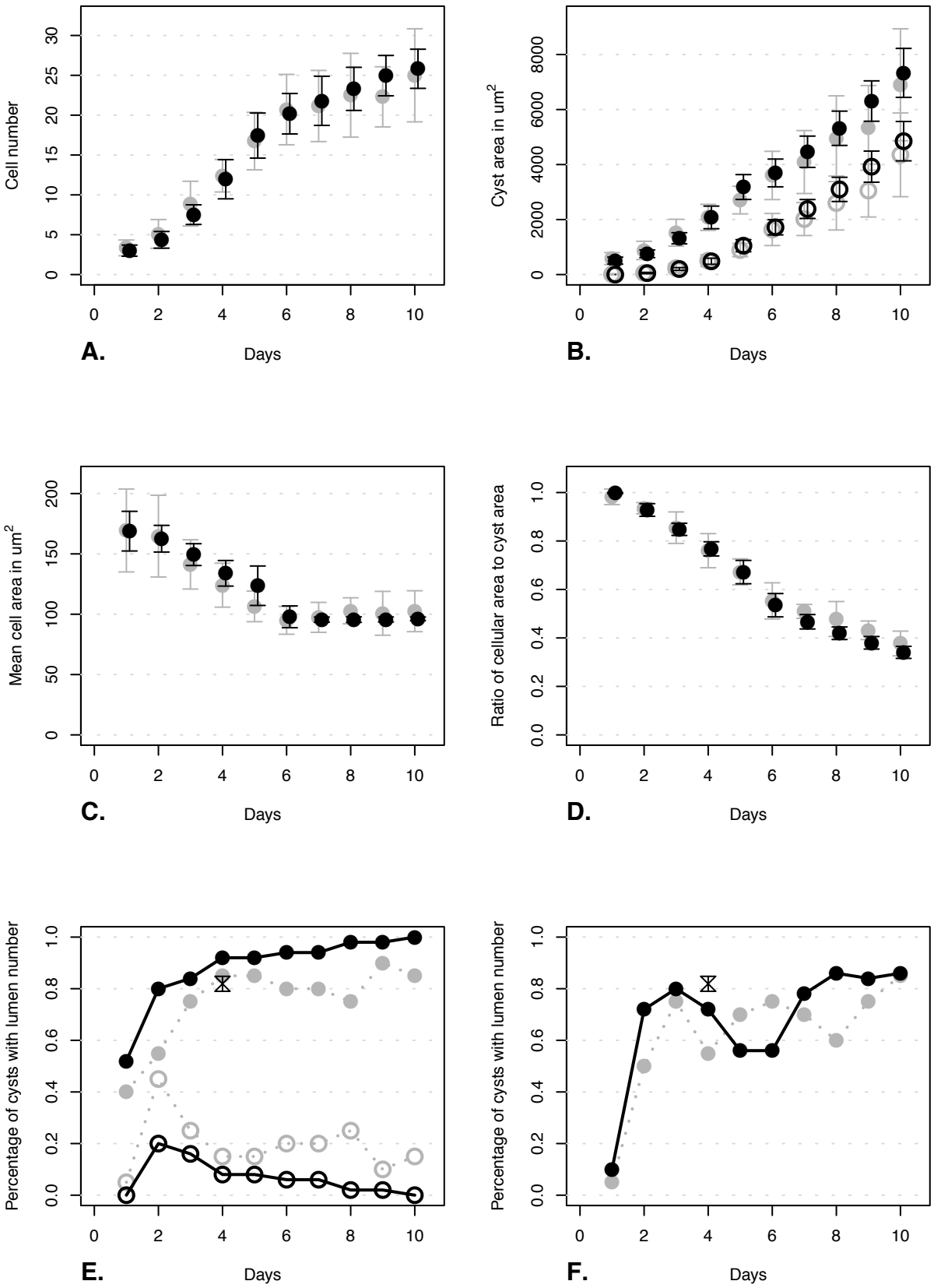
Appendix 1, Figure 64. Varied deathRateLumen: drl = 0.1



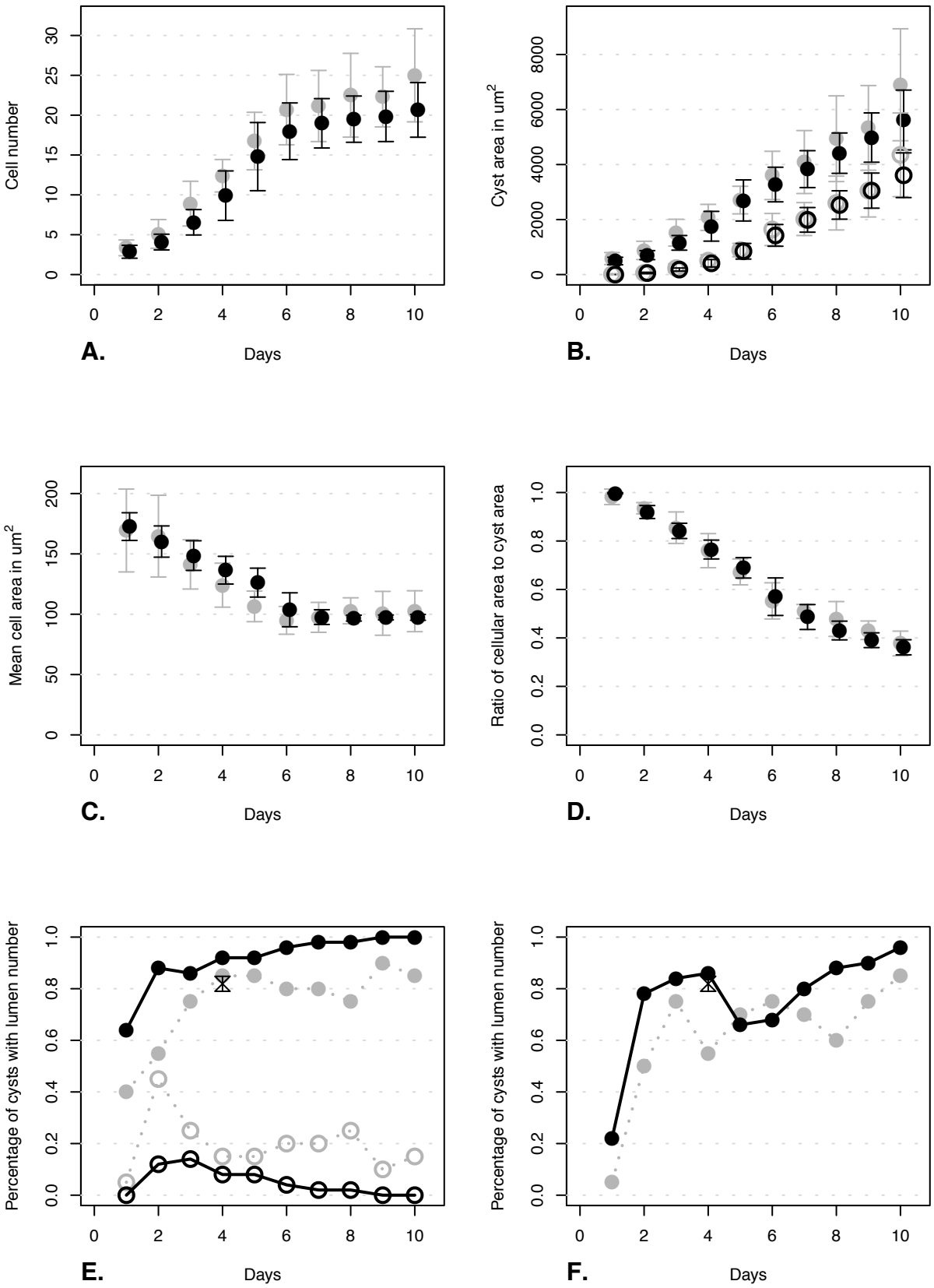
Appendix 1, Figure 65. Varied deathRateEpi: dre = 0.0



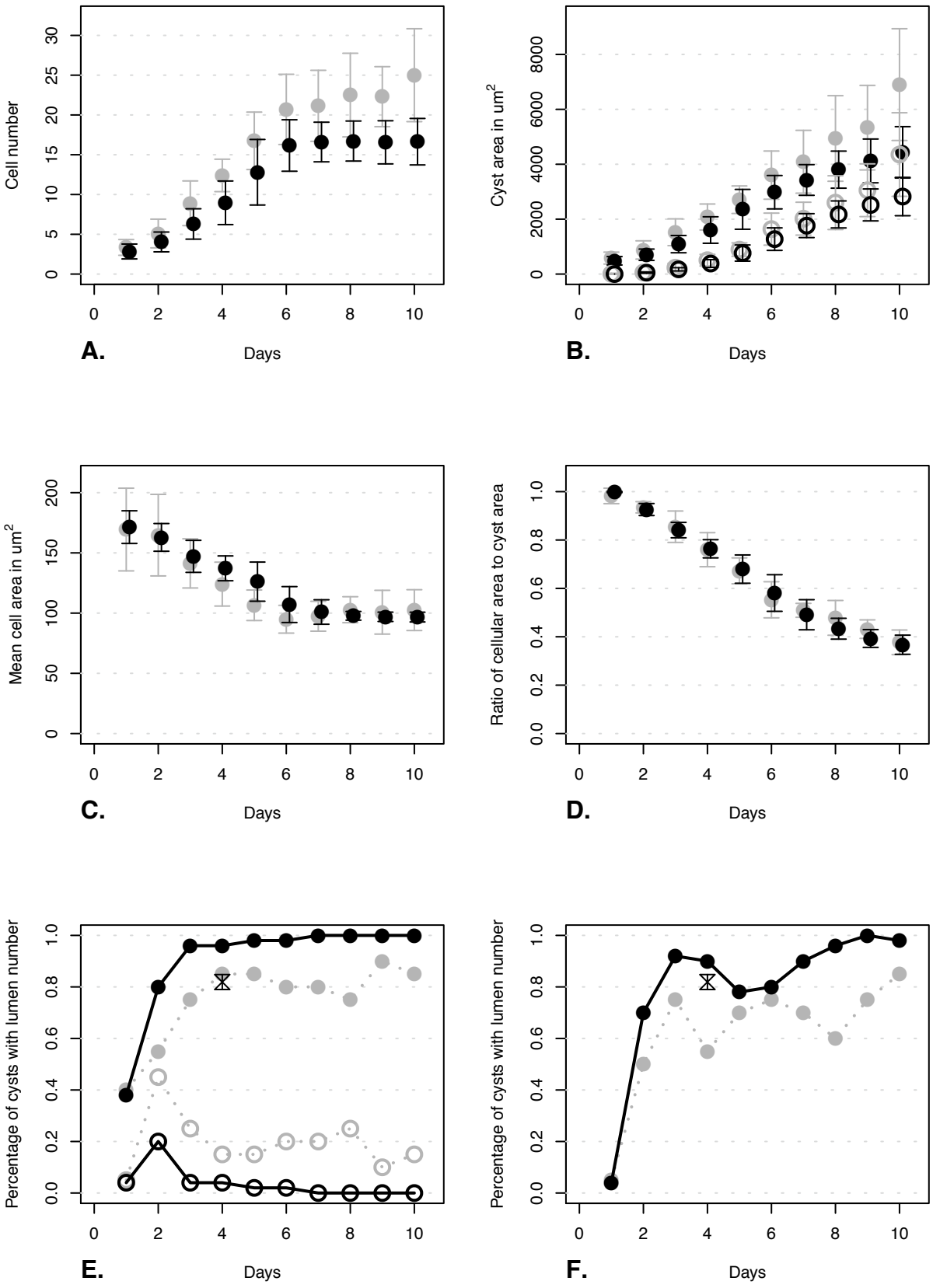
Appendix 1, Figure 66. Varied deathRateEpi: dre = 0.0002



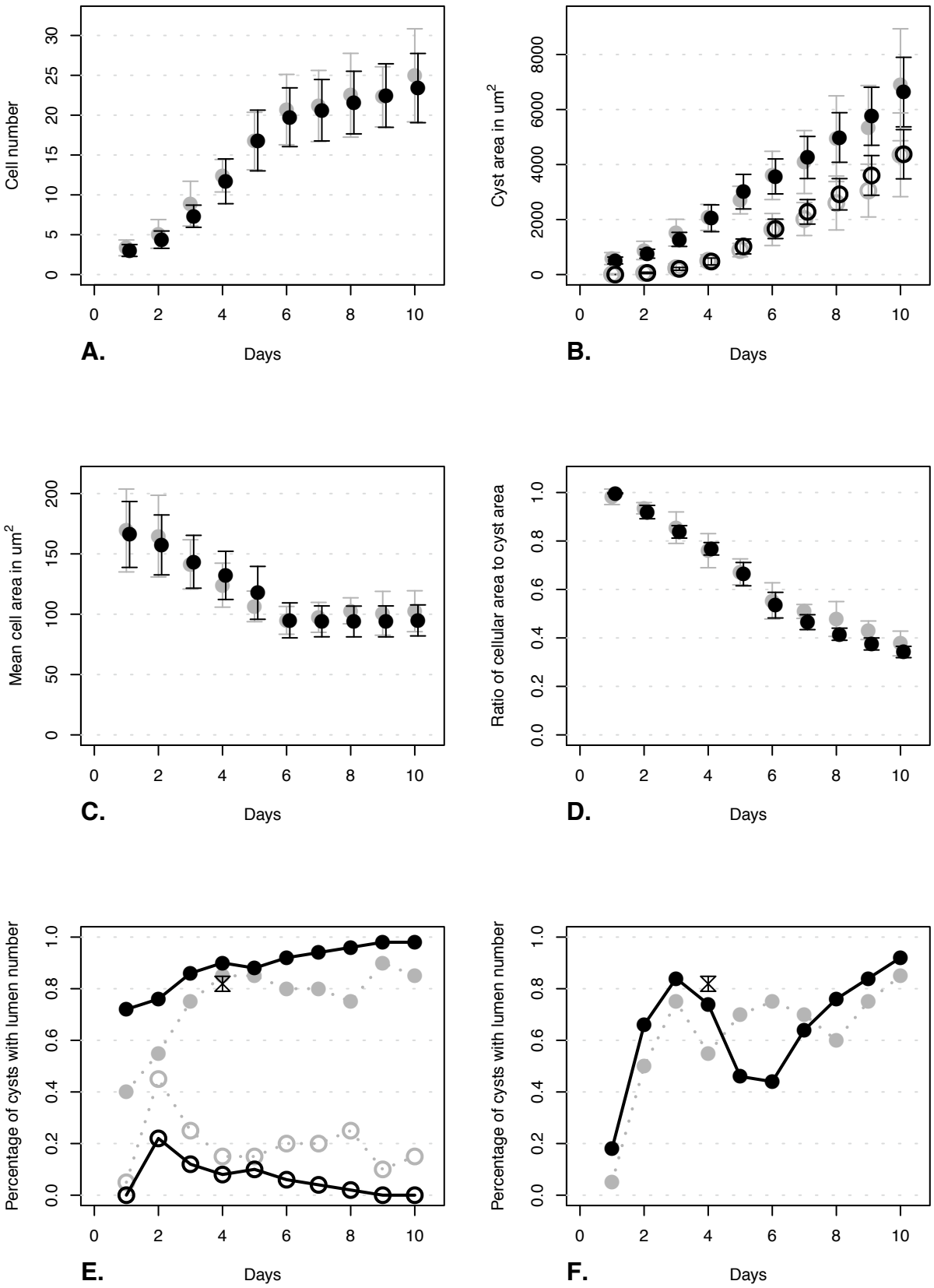
Appendix 1, Figure 67. Varied deathRateEpi: dre = 0.0008



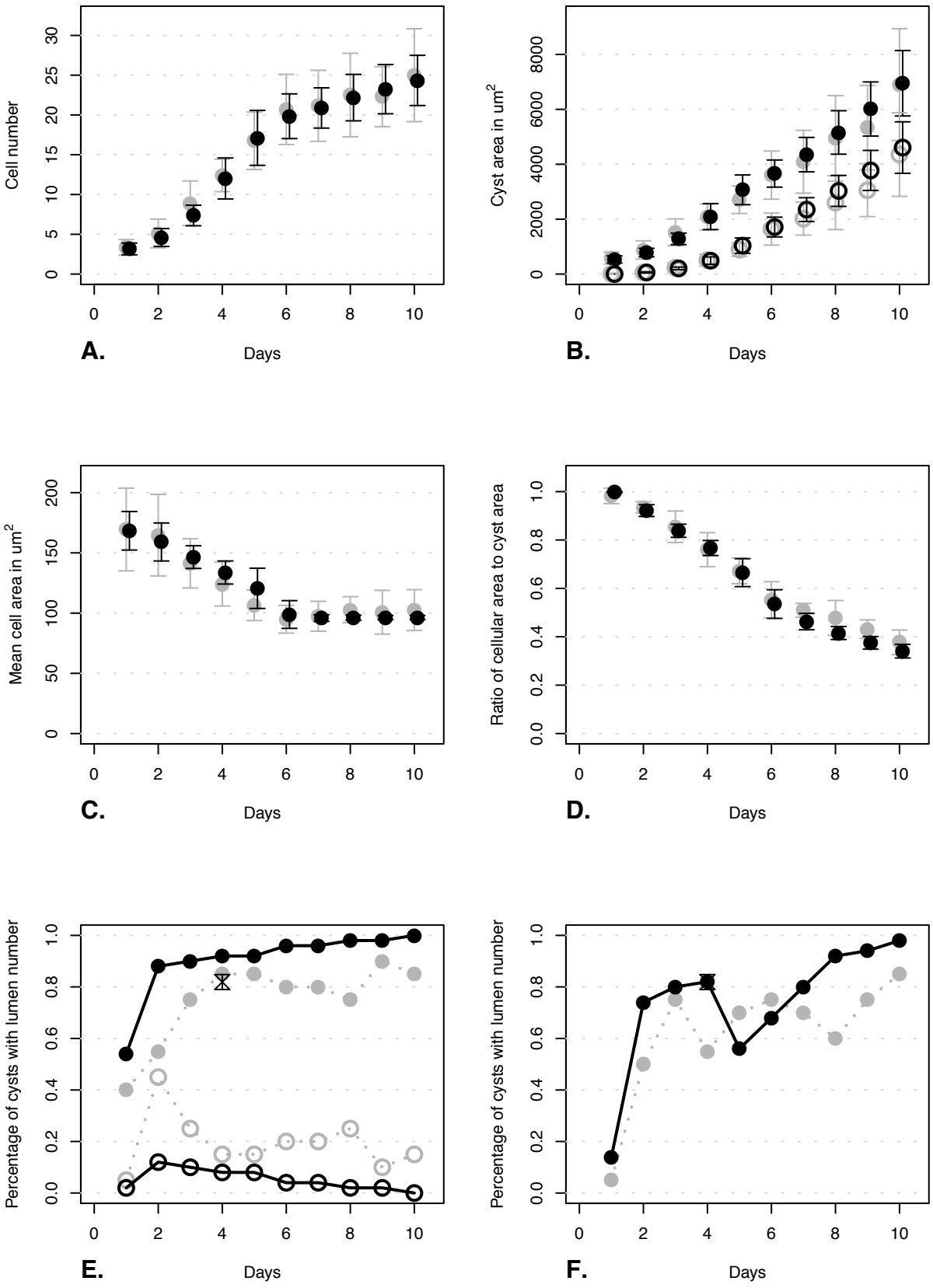
Appendix 1, Figure 68. Varied deathRateEpi: dre = 0.0016



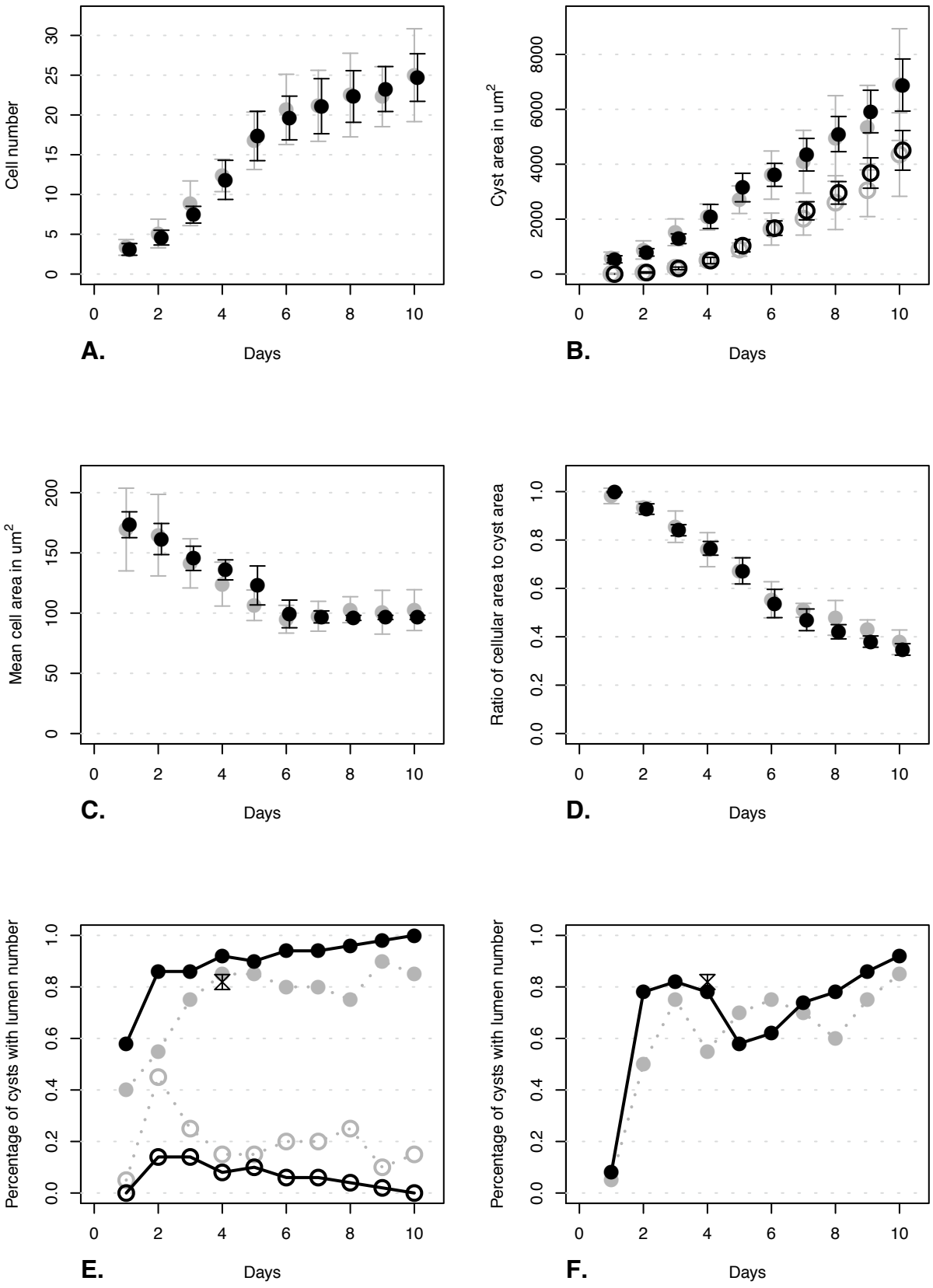
Appendix 1, Figure 69. Varied dyingShrinkRate: dsr = 1.75



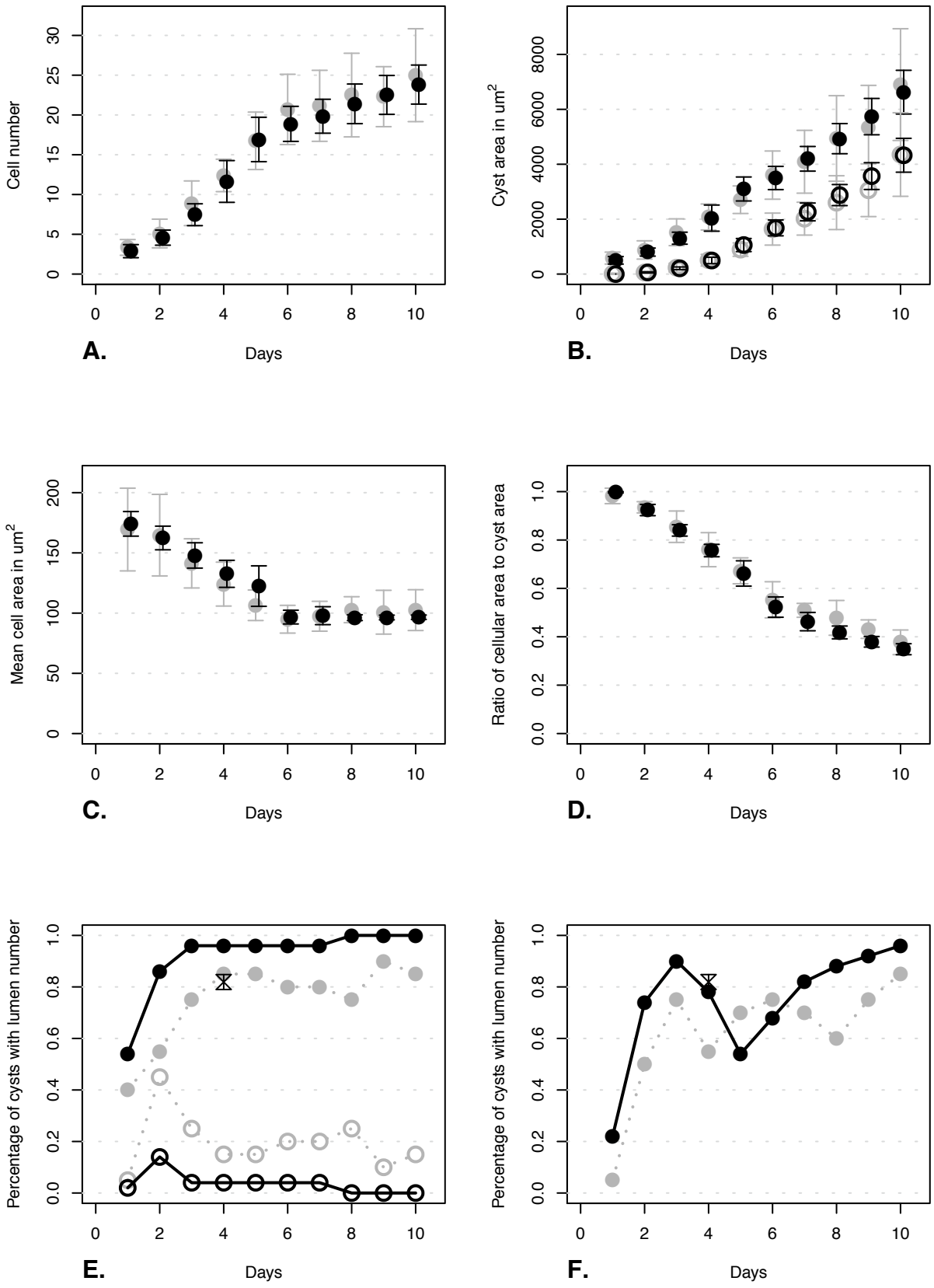
Appendix 1, Figure 70. Varied dyingShrinkRate: dsr = 4.5



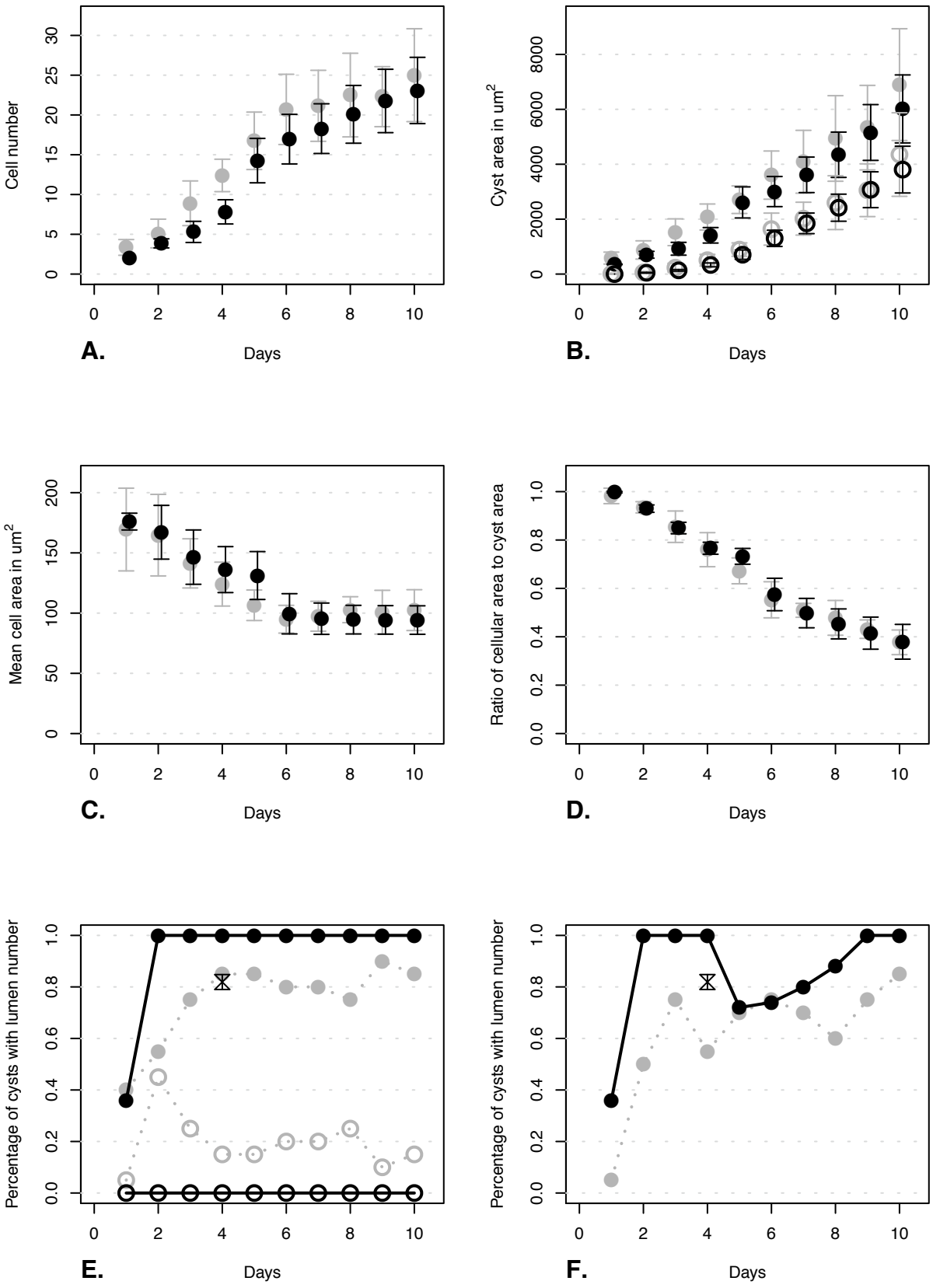
Appendix 1, Figure 71. Varied dyingShrinkRate: dsr = 13.5



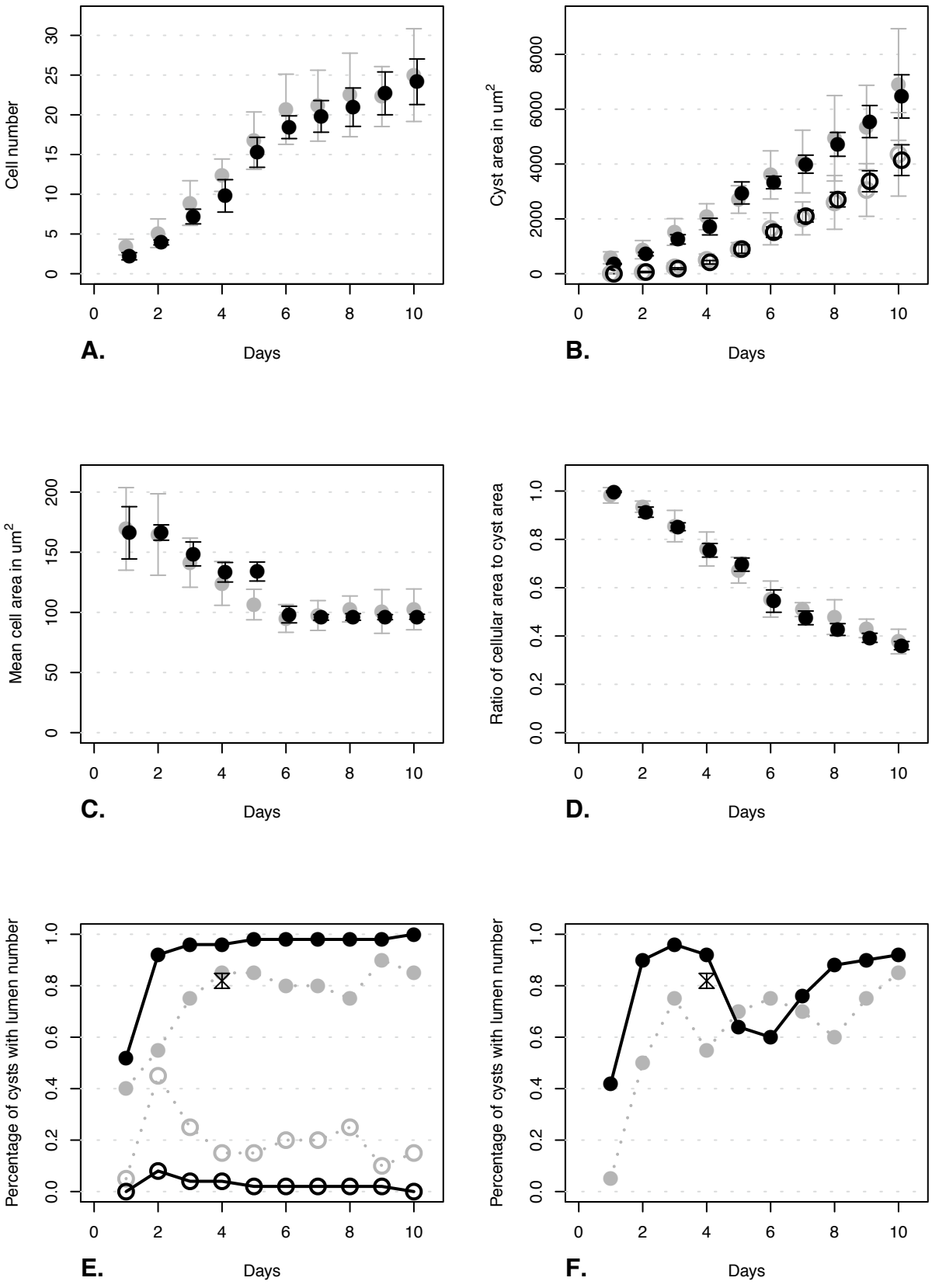
Appendix 1, Figure 72. Varied dyingShrinkRate: dsr = 18



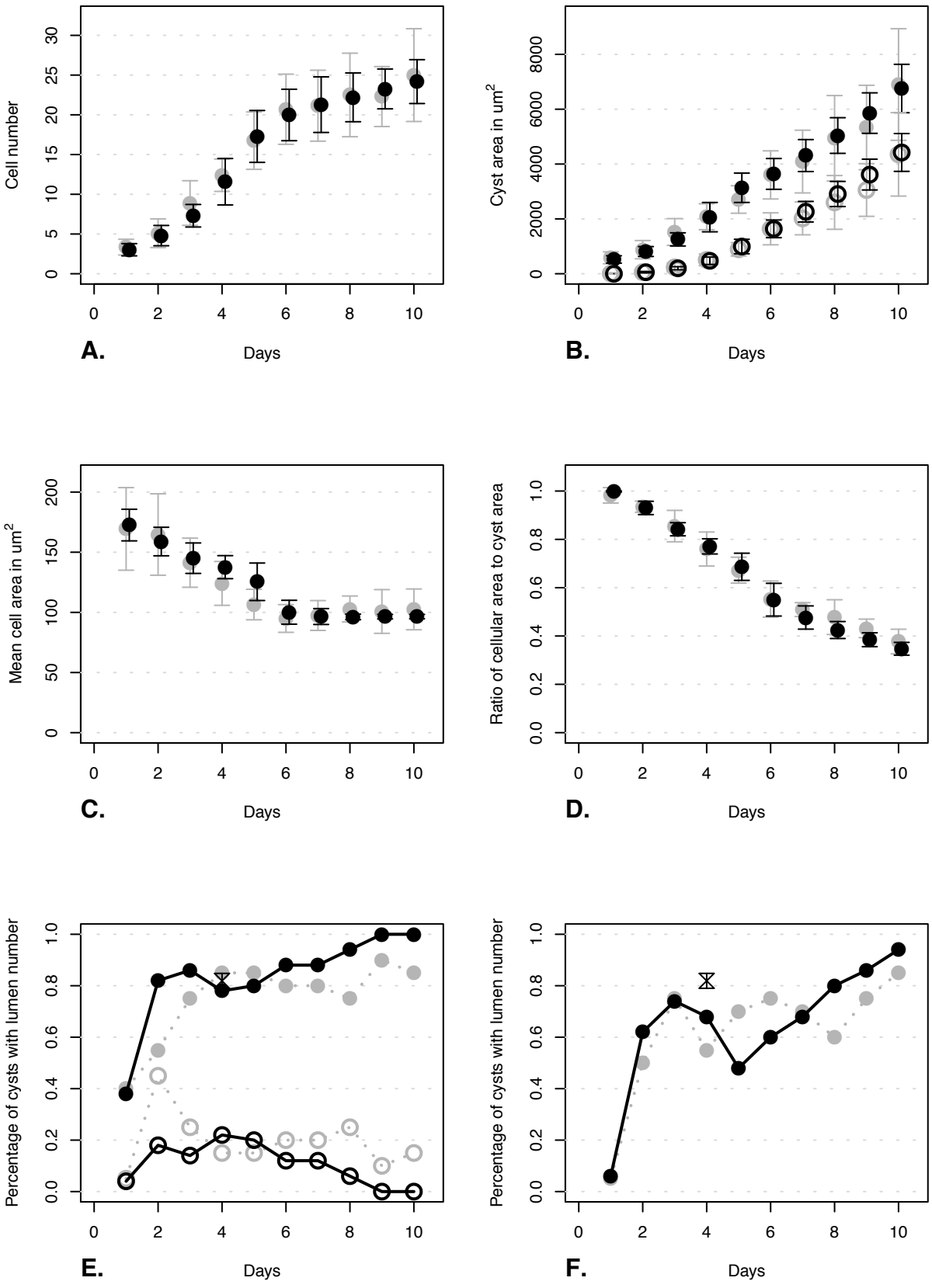
Appendix 1, Figure 73. Varied clusterProb: cp = 0



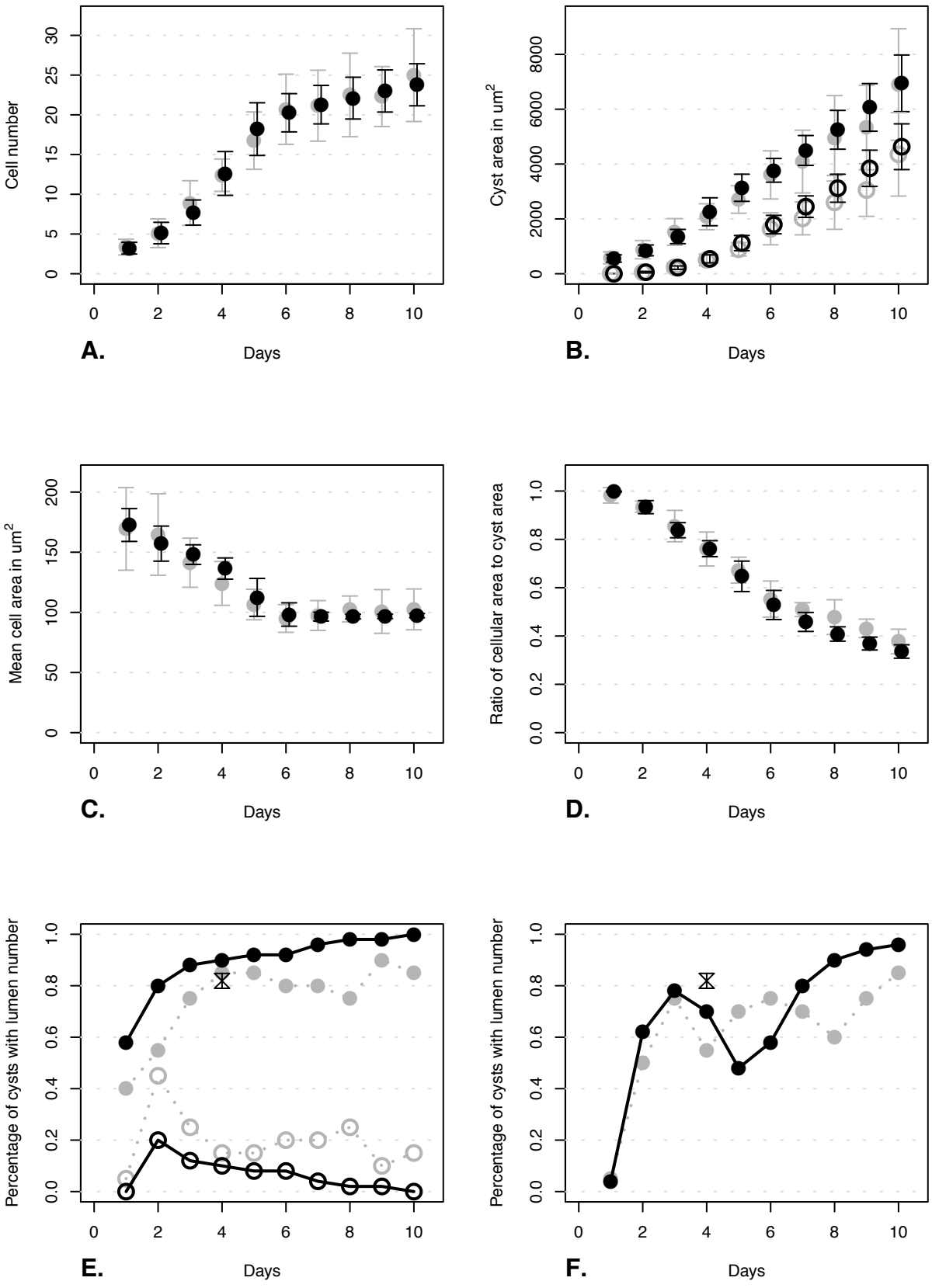
Appendix 1, Figure 74. Varied clusterProb: cp = 0.4



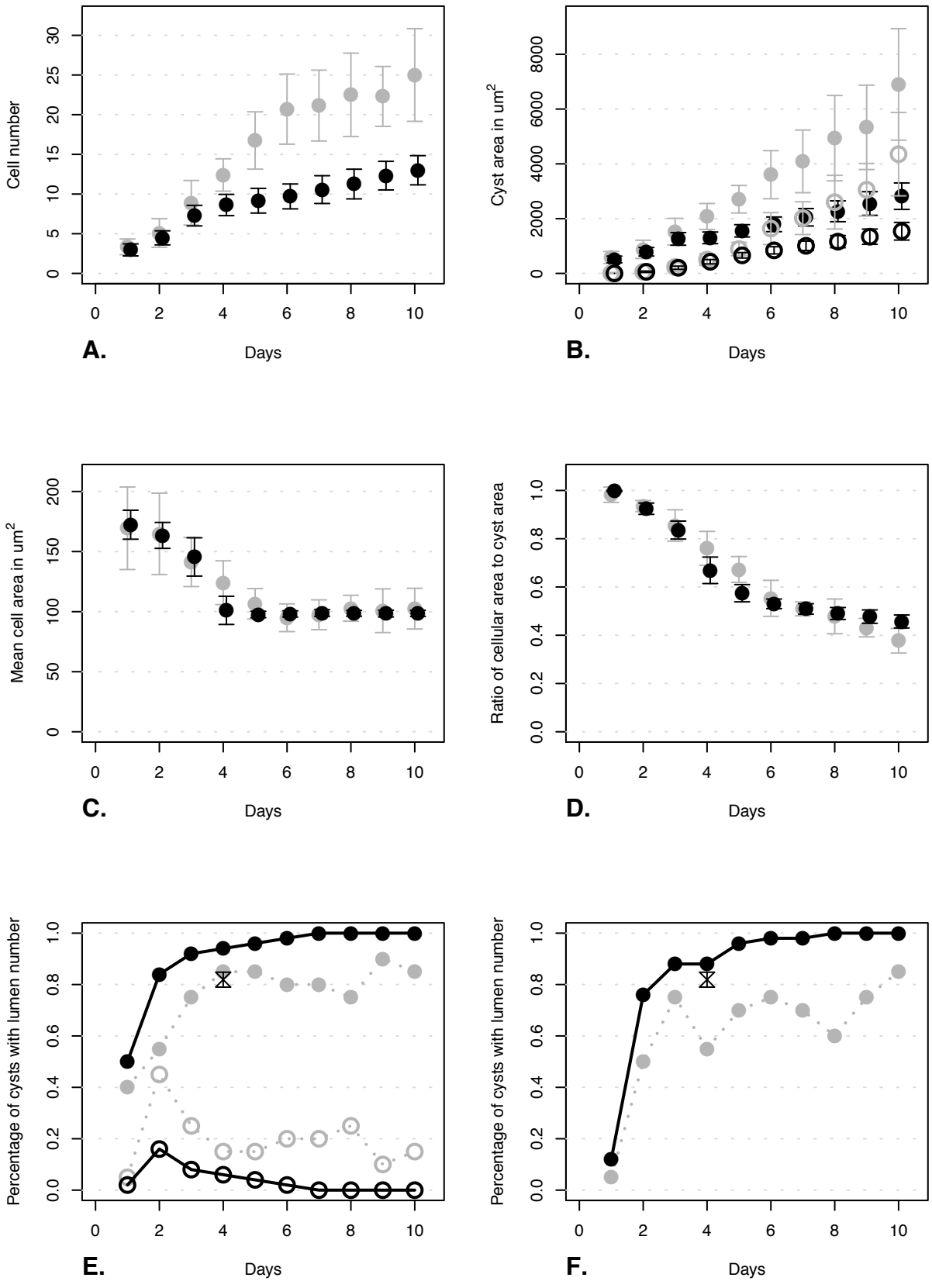
Appendix 1, Figure 75. Varied clusterProb: cp = 0.9



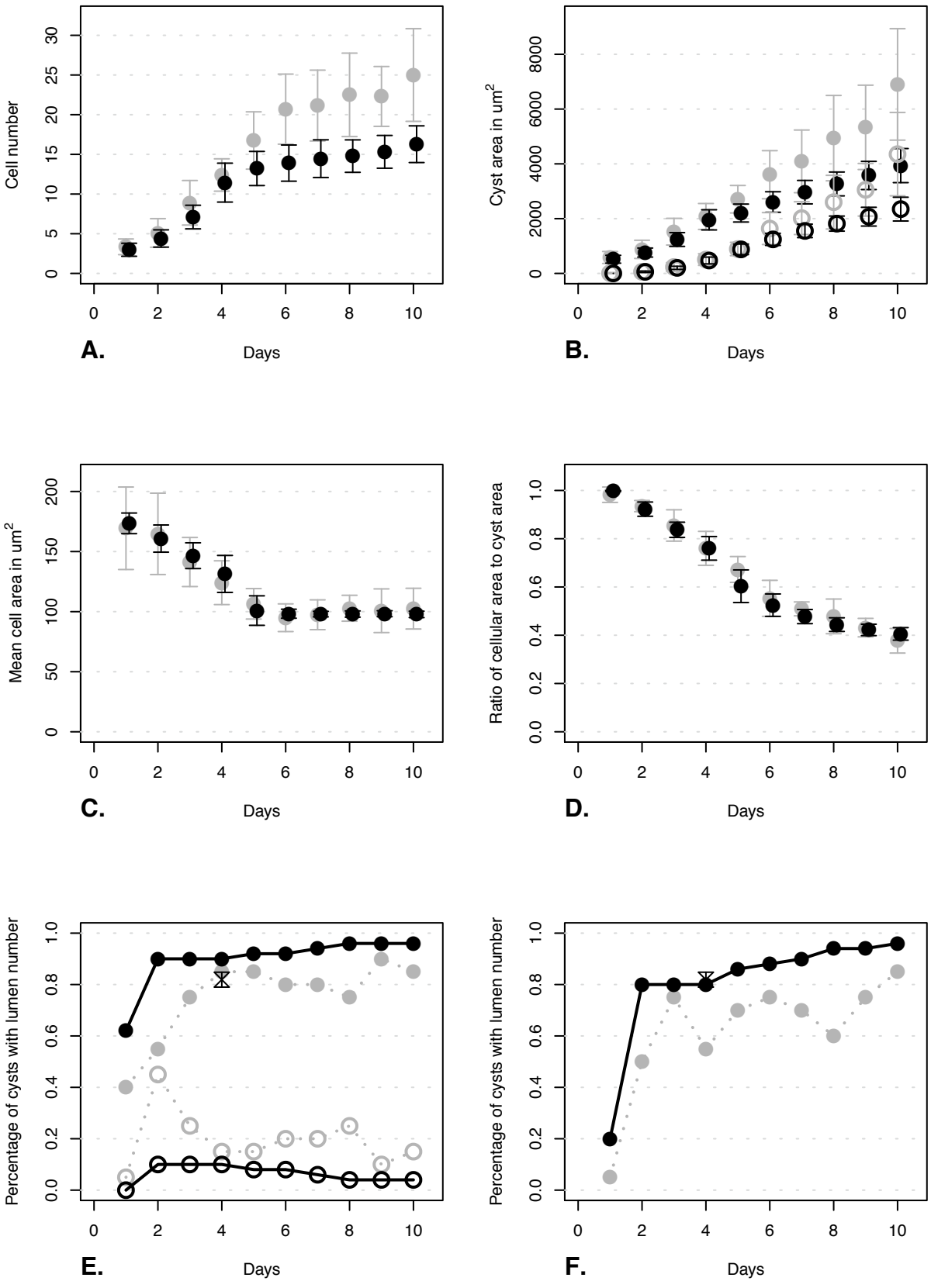
Appendix 1, Figure 76. Varied clusterProb: cp = 1



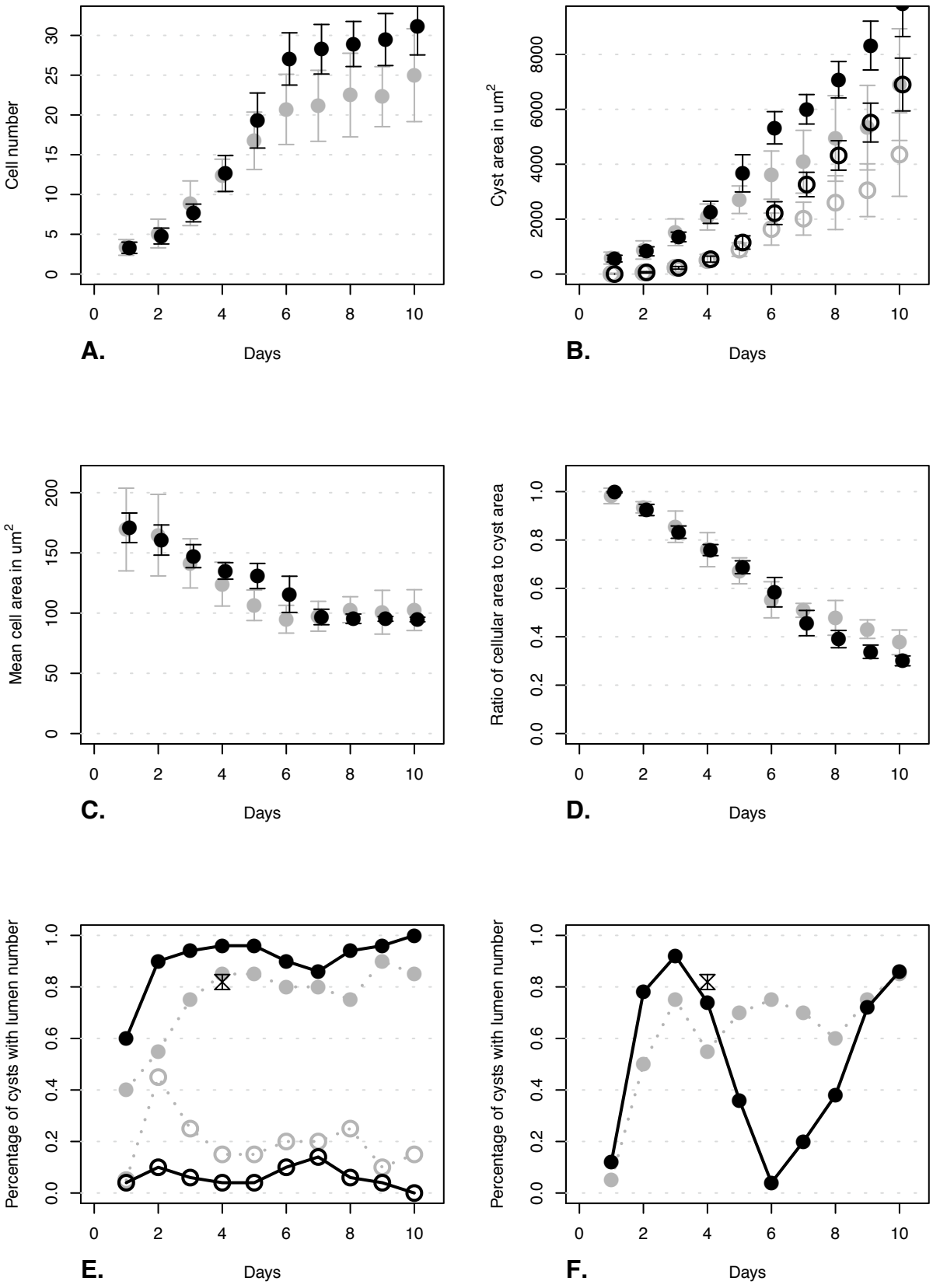
Appendix 1, Figure 77. Varied stableRatio: sr = 0.125



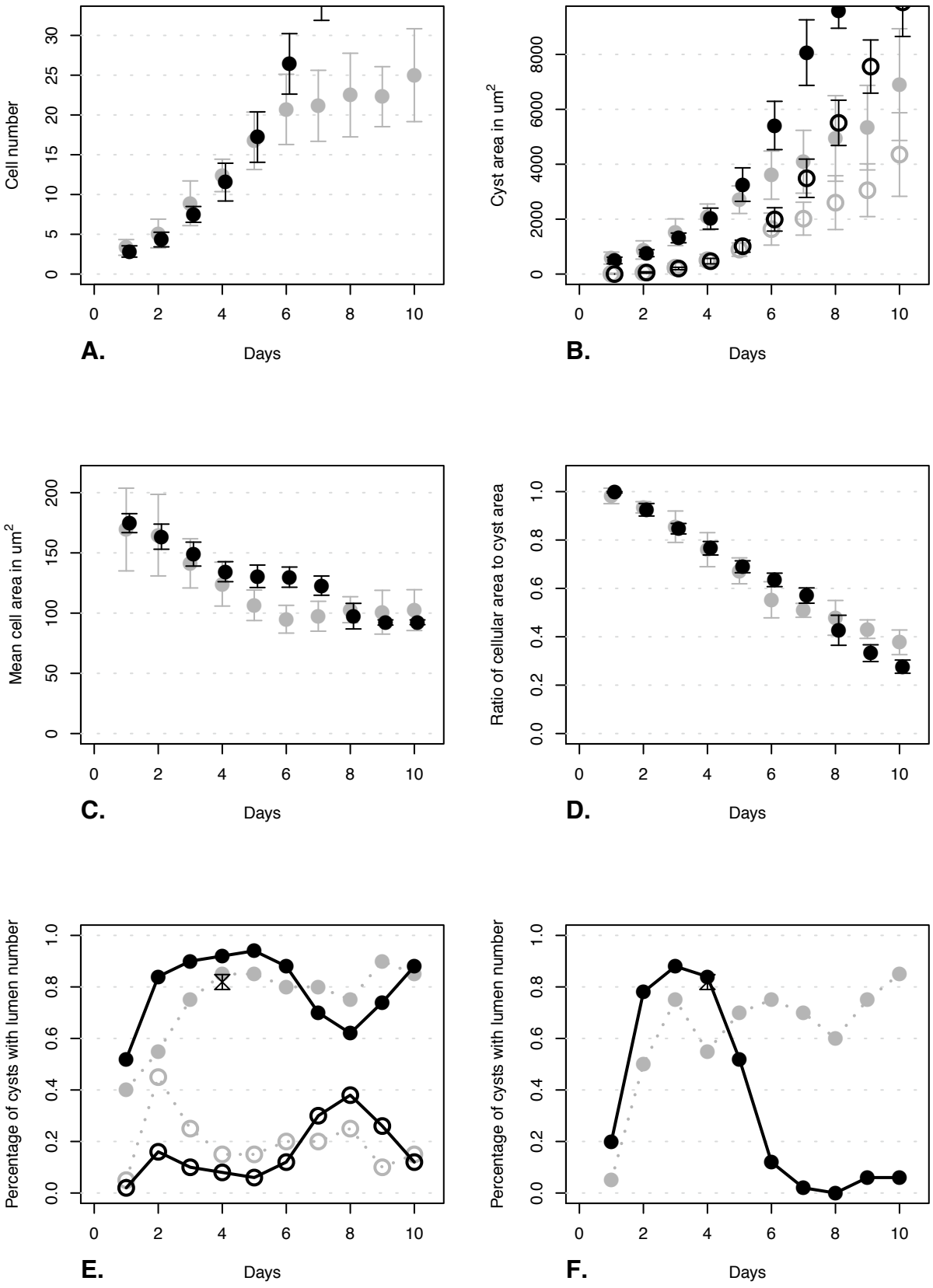
Appendix 1, Figure 78. Varied stableRatio: sr = 0.25



Appendix 1, Figure 79. Varied stableRatio: sr = 1.0



Appendix 1, Figure 80. Varied stableRatio: sr = 2.0



8. Appendix 2.

A Robust in Silico Analogue of MDCK Cystogenesis Mimics Growth in Multiple Culture Conditions

Jesse A. Engelberg[1], Anirban Datta[2], Keith E. Mostov[3], and C. Anthony Hunt[1,3]

[1] **Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California 94143**
[2] **Department of Anatomy, University of California, San Francisco, California, 94143**
[3] **Joint Graduate Group in Bioengineering, University of California, San Francisco and Berkeley, California 94143**

jesse.engelberg@gmail.com a.hunt@ucsf.edu

Keywords: cell biology, composite agent, cross-model validation, morphogenesis, synthetic modeling, systems biology

Abstract

Madin-Darby canine kidney (MDCK) cells undergoing cystogenesis in vitro is a scientifically useful model of epithelial morphogenesis. The cysts formed in collagen and Matrigel are qualitatively similar, consisting of a single layer of epithelial cells surrounding a hollow lumen. However, differences in key quantitative measures of cyst growth, including cell number and cyst and lumen size, indicate that some cell behaviors are different within the two culture systems. We recently described an agent-oriented, agent-directed analogue of MDCK cystogenesis in Matrigel. It utilized a cellular Potts model and achieved qualitative and quantitative validation targets using empirical parameter tuning. Within this report we highlight steps taken to convert the cellular Potts model framework to one based upon an agent-oriented approach. If measures of cell death are ignored, the only parameters that required adjustment to allow the analogue of cystogenesis in Matrigel to mimic MDCK cystogenesis in collagen were those controlling cell division and polarization. These data indicate that in addition to delayed cell polarization, cell division in collagen is likely slower than in Matrigel. The reported results support the hypothesis that MDCK cells use the same basic operating principles to create cysts when cultured in Matrigel or collagen.

1. INTRODUCTION

In vitro cystogenesis by MDCK cells is a useful model of epithelial morphogenesis and organogenesis. Growth of cysts in culture mimics many behaviors known to occur during epithelial organ development. When MDCK cells are cultured in Matrigel or collagen their behavior follows the same qualitative pattern, with cell division leading to cell polarization, lumen creation, and lumen expansion [1]. However, some

features, including cyst size and the onset of lumen formation, are obviously different, as documented in [1]. Available evidence from in vitro experimentation does not indicate whether the differences are the result of cells following similar operating principles within different environments, or whether events within collagen cultures are a consequence of operating principles having one or more fundamentally different features. The development, study, and challenge of agent-oriented analogues of in vitro systems are a new approach to achieving an improved understanding of cell biology. Once properly validated, such analogues provide valuable new insight into mechanisms that may be responsible for referent system behaviors. An in silico MDCK analogue of cell growth within Matrigel cultures (called ISMA-M) was recently reported [2], along with novel quantitative data used for validation. The focus of this report is new in vitro data for MDCK cystogenesis in collagen. The ISMA-M was iteratively refined so that simulation results mimicked most of that data: the resulting analogue is called an ISMA-C. Only parameters influencing cell number and the timing of lumen initiation were altered, yet the new ISMA-C survived strong falsification challenges. Measures of events during simulations provided values for ISMA-C cell number, cyst and lumen size, mean cell area, the ratio of cellular to cyst area, and lumen number percentages that were acceptably similar to those observed in vitro. Within ISMA-C, cyst and lumen size are linked to cell number and onset of lumen formation. We hypothesize that the same is true in collagen cultures. This paper demonstrates that a single in silico analogue, when parameterized differently, can mimic in vitro results generated using different culture systems. It also serves to illustrate the usefulness of the Iterative Refinement Protocol [2-5] and sets the stage for future modeling and simulation efforts.

2. METHODS

Hereafter, to avoid confusing wet-lab with simulated features and behaviors, we use SMALL CAPS when referring to the latter. To generate quantitative data for ISMA-C validation, MDCK cells were grown suspended within collagen (2 mg/ml) for ten days. Daily samples were fixed and imaged using confocal microscopy, as done previously for Matrigel cultures [2]. Although a higher cell density was used for collagen relative to Matrigel cultures, the frequency of cell clustering was lower. That was because clusters and cysts in Matrigel form on top of a layer of 100% Matrigel as opposed to being fully suspended, as was the case in collagen.

The detailed descriptions in Engelberg et al. [2] for the ISMA system, parameters, and computational methods apply here. A considerably abridged description follows. The ISMA-C was constructed so that components and mechanisms map to in vitro counterparts. The system included CELLS, LUMINAL space, and EXTRACELLULAR MATRIX. Components existed on a 2D hexagonal grid in which CELLS and LUMEN occupied multiple grid locations with the remaining space classified as MATRIX (Fig. 1). ISMA-C was implemented using the CompuCell3D [3] cellular Potts model framework and custom code (as described below). Individual CELLS could expand, DIVIDE, change shape, create LUMENS, move, or change state depending on their internal variables and external environment (Fig. 2). CELLS applied operating principles from specifications developed from a set of targeted in vitro MDCK cystogenesis attributes. Some of these

operating principles, such as CELL DIVISION, were modeled on the mechanisms observed within in vitro culture.

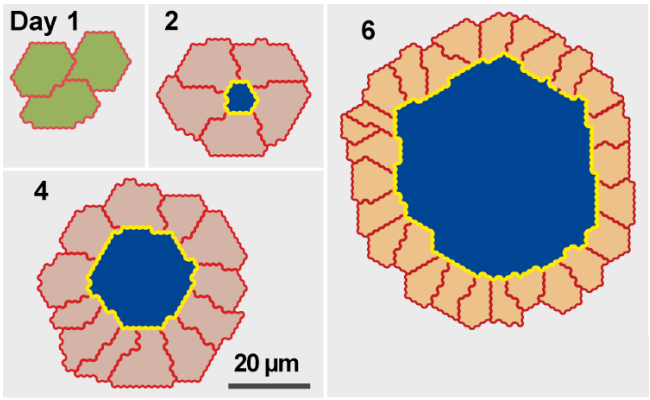


Figure 1. In silico MDCK analogue CYST cross sections. Note that a regular hexagon in hexagonal space maps to a circle in continuous space. Images are from a single simulation run using parameter settings for growth in simulated Matrigel. Growth in simulated collagen produces similar qualitative results. CELLS are UNPOLARIZED (green), POLARIZED (gray) or stabilized (orange). CELL-CELL and CELL-MATRIX borders are red; CELL-LUMEN borders are yellow; LUMENS are blue.

An Iterative Refinement Protocol [4-7] was used to adjust CELL behaviors. The IR Protocol is as follows: select a set of targeted attributes that describe behaviors of the referent in vitro system; specify and then implement an analogue to mimic these targeted attributes; falsify the analogue using Similarity Measures (SMs) based on in vitro data specific to the selected targeted attributes; modify the analogue code and/or parameters until it survives falsification. When this is complete, the analogue is validated for that set of targeted attributes. To increase the behavior space captured by the analogue the IR Process is continued; expand the list of targeted attributes so that the current analogue is falsified by SMs based on in vitro data specific to the new targeted attribute; refine the code and/or parameters until the analogue survives falsification. The new collagen data provided a strong challenge to ISMA-M, which was falsified when using the parameter settings in [2]. To obtain an ISMA-C that would survive falsification, ISMA-M parameter settings were iteratively modified until a satisfactory match to in vitro data was achieved.

In order to mimic the observed smaller cell numbers during the first four days, *clusterProb* was reduced to 0.7 from 0.8 and *cellCycle* was increased from 42 to 100 (the latter maps to 50 hours). Following that refinement the ISMA-C achieved some similarity targets, but LUMENS formed too early. The value of *polarDelay* was adjusted until LUMEN formation in silico matched that in vitro. Surprisingly, no additional changes to parameter settings were required. The underlying code was not modified to produce these results.

The cellular Potts model is a modified cellular automata containing one or more CELLS, each occupying multiple locations on a grid, allowing for the simulation of CELLS with varied size and shape [8, 9]. During a simulation cycle, individual locations on the borders of a CELL can become contained to a neighboring CELL according to probabilistic rules. One rule is that each CELL has an ideal size (number of grid loca-

tions), and any grid location change that brings the CELL closer to that ideal is favorable. The CompuCell3D framework implements the CPM, and allows users to simulate biological processes, automating things such as CELL size and shape change, CELL ADHESION, CELL DIVISION, and graphical visualization.

The ISMAs used agent-oriented, agent-directed methods. The system contained CELL and LUMEN agents, as well as an agent that executed the underlying cellular Potts model (cellular Potts models are detailed in [8, 9]). CELL DIVISION and CELL state change occurred during the execution of CELL agents, while changes to individual grid locations occurred during the execution of the Potts agent. During the latter process, a random set of grid locations was evaluated to see if their index should change from being contained within a specified CELL to being contained within a neighboring CELL or LUMEN. That protocol controlled behaviors such as CELL size and shape changes, LUMEN expansion, and CELL shrinkage after initiating DEATH. To calculate whether a point would change its index, the Potts agent evaluated the resulting energy change, weighing the changes in CELL area and perimeter, the change in configuration of CELLULAR neighbors, and whether CELLS would be separated or internal rules violated. The latter resulted in an energy penalty. Special rules governing LUMEN expansion prevented CELLS from coming into contact with multiple LUMENS. Each CELL had ideal area and perimeter targets. Changes that created deviations from those targets resulted in higher energies and thus were less likely to occur.

CELLS executed operating principles that dictated when they would change state, DIVIDE, create a LUMEN, or DIE. CELLS had three states: UNPOLARIZED, POLARIZED, and stable. CELL POLARIZATION and DIVISION occurred after a set amount of time elapsed, while LUMEN creation occurred when a POLARIZED CELL contacted other CELLS and the MATRIX, but did not contact a LUMEN. CELLS stabilized when they sensed that their neighboring LUMEN had reached a critical size, controlled by the parameter *stableRatio*. CELL DEATH occurred at random, with CELLS being more likely to DIE when not in contact with the MATRIX.

The stock implementation of CompuCell3D was designed using a grid-based and system-based approach. Individual CELL state change was often initiated after index change events during the Potts execution step. CELLS were not aware of the points contained within them, and there was no way for an individual CELL to execute DIVISION directly. In order to make the process more biomimetic, we developed a custom plug-in that allowed the simulation to be executed from the perspective of individual CELLS. When this plug-in was called, it stepped through each CELL and allowed it to evaluate its internal state and external environment, changing variables such as the target area or polarization state as a result. In addition an *MCell* object was added to each CELL to create a bidirectional mapping between CELLS and grid points. Each CELL was mapped to an *MCell*, which contained references to the grid points contained within that CELL. These modifications allowed the analogue to execute using a perspective that was more intuitive, and that mimicked understanding of intracellular biology, where the cell acts as the fundamental functional unit. This made modifications significantly easier to implement. For example, CELL DIVISION was implemented in an agent-based manner so CELLS had access to information about the points contained within them, thus it was straightforward to randomize or invert the axis of CELL DIVISION.

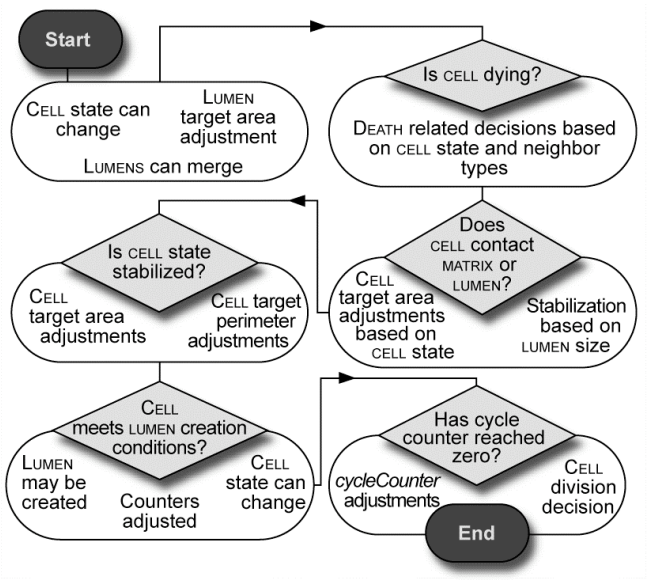


Figure 2. ISMA logic and decision control flow. CELLS step through five logic modules during a simulation cycle, executing actions according to their current state and external environment. Actions include the adjustment of LUMEN target area, CELL DEATH, adjustment of CELL area and state, creation of new LUMENS, and the creation of new CELLS.

In order to validate the quantitative results of ISMA-M and ISMA-C, similarity measures were developed for cell number, lumen and cyst size, mean cell area, and the ratio of cellular to cyst area [2]. Similarity Measure 1 (SM1) measured the percentage of in silico simulations at a given day that were within ± 25% of the mean in vitro value at that day, thus finding a measure of similarity between in silico and in vitro observations. Based on knowledge of the in vitro system and previous experience designing similarity measures, we judged that if 50% or more simulations fell within that 25% of the in vitro mean (and thus the SM1 value was > 0.5) for 8 out of 10 days, the simulation survived falsification for SM1. The variance of in vitro results can have a significant effect on how stringent the SM1 cutoff should be, so to measure in vitro variance in a way that could be directly compared to SM1, Self-Similarity Measure 1 (SSM1) was developed. SSM1 was exactly the same as SM1, except it measured the similarity between individual in vitro runs and the mean value generated by these runs. Observations with high values of SSM1 had high variance. To directly compare the variance of in silico and in vitro results, rather than simply using the in vitro variance as a guide, Similarity Measure 2 (SM2) was developed. SM2 equaled the absolute value of the difference between the coefficients of variation of the in vitro data and the in silico data at each day. A smaller value of SM2 at a given day indicated that the coefficient of variation was similar for in vitro and in silico data. A simulation survived falsification for SM2 if the value of SM2 was less than 0.25 for 9 out of 10 days for a given measure.

Table 1. *Glossary*

Biological terminology
Collagen. A cell culture medium consisting of the collagen I extracellular matrix protein.
Cystogenesis. The process by which cells divide and organize into a hollow cyst consisting of a single (usually) layer of cells.
Epithelial cells. Cells that make up epithelial organs (such as the skin, lung, kidney), provide a barrier function, and are often polarized, expressing different proteins on either side.
Extracellular matrix (ECM). The protein component surrounding cells that acts as a scaffold for cellular structure.
Lumen Any hollow space, often inside a cyst or duct.
Madin-Darby canine kidney (MDCK) cell. A particular type of dog kidney cell used in cell culture experiments. Capable of surviving in cell culture but lacking most cancerous behaviors.
Matrigel. A complex cell culture medium secreted by Engelbreth-Holm-Swarm (EHC) mouse sarcoma cells containing a number of extracellular matrix proteins.
Morphogenesis. The general process of cellular organization into simple or complex structures.
Polarization. The process or state in which cells express varied proteins in opposite sides, while adjacent sides are in contact with neighboring cells.
Modeling terminology
Biomimetic. Something that imitates biology.
Cellular Potts model (CPM). A cellular automata in which CELLS can occupy multiple locations to model cell size and shape change.
CompuCell3D. A CPM-based framework that automates cell size change and visualization.
ISMA. In Silico MDCK Analogue
ISMA-M. An ISMA parameterized to mimic MDCK cystogenesis within Matrigel culture.
ISMA-C. An ISMA parameterized to mimic MDCK cystogenesis within collagen culture.
MCell. An ISMA object that allows CELLS to access the points contained within them.
Operating principle. A rule or set of rules that translates internal state and environmental cues into observable actions.
Iterative Refinement (IR) process. The process by which the ISMA is refined and improved over time.
Similarity Measure. A numerical measure of the similarity between in vitro and in silico data.
Stabilized. An in silico CELL state in which CELL DIVISION is slowed and other operating principles change. Hypothesized to have an in vitro equivalent.

3. EXPERIMENTS AND RESULTS

Cyst growth within collagen was similar to that observed during previous experiments [1]. Suspended cells divided to form clusters, polarized after two to three days, and formed single or multiple lumens by day 3. Lumens and cysts expanded in size thereafter. Within Matrigel cultures, a change in the rate of cell division and cell size was observed at day 6. A similar change was observed within collagen cultures. Growth slowed at roughly day 6, and mean cell size, which had decreased between days 2 and 6, leveled off. After empirical parameter turning, ISMA-Cs successfully mimicked those qualitative behaviors. Measures of ISMA-C simulations also mimicked quantitative measures of growth within collagen. Prespecified similar-

ity measures (SM1 and SM2) were achieved for CELL number, CYST size, mean CELL area, and ratio of CELLULAR to CYST area, but not for LUMEN size. In order to achieve similarity for CELL number, the parameters *clusterProb* and *cellCycle* were altered so that in silico growth matched in vitro growth during days 1-5 (Fig. 3A). ISMA-C simulations produced CELL growth numbers similar to in vitro during days 6-10 without additional tuning. However, CELL stabilization occurred at day 8 instead of at day 6 as observed in vitro.

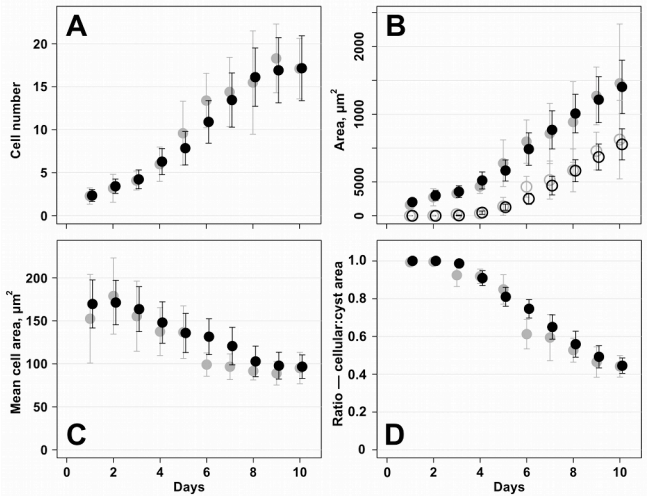


Figure 3. Quantitative values for in vitro and in silico cystogenesis in collagen. Mean values and standard deviations for (A) cell number per cyst, (B) cyst (solid circles) and lumen (open circles) area, (C) mean individual cell area and (D) ratio: cellular to cyst area. Gray: in vitro data taken each day for ten days from 10 cysts. Black: data taken from 50 CYSTS over ten DAYS using specified

Cells within collagen polarize and initiate lumens somewhat slower than did their counterparts within Matrigel (Fig. 3B). In order to mimic that behavior, the variable *polarDelay* (which controls the delay before CELLS change to the POLARIZED state) was empirically tuned until mean LUMEN formation time was similar to that observed in vitro. Doing so required increasing *polarDelay* from 42 (which maps to 21 hours) to 120 simulation cycles (which maps to 60 hours). With that parameter setting, simulations produced mean CYST and LUMEN sizes that were remarkably similar to those observed in vitro (Fig. 3B). It is important to note that variables controlling LUMEN expansion, the effect of CELL stretching on LUMEN expansion, and the mean size of CELLS were not altered, yet ISMA-C results closely mimicked those observed in vitro, indicating that *polarDelay* was the primary driver between these varied simulation results.

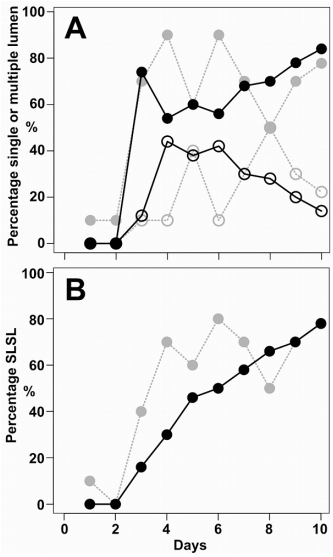


Figure 4. Percentage of cysts with specified types of lumens. (A) Percentage of cysts with single (solid circle) or multiple (open circle) lumens. (B) Percentage of cyst with a single lumen surrounded by a single layer (SLSL) of cells. Gray: in vitro data for 10 cysts taken each day for ten days. Black: in silico data for 50 CYSTS using specified parameter settings. Solid lines: continuous ISMA-C CYST growth. Dotted lines: discrete MDCK cyst growth in collagen.

Mean cell size within collagen did not differ significantly from that within Matrigel. The good quantitative match between mean ISMA-C and in vitro cell sizes (Fig. 3C) was evidence that the ISMA-C analogue was functioning as intended. Lumen size within in vitro Matrigel and collagen culture differed significantly, especially at later stages of growth. It was thus encouraging to observe that the ratio of CELLULAR to CYST areas during ISMA-C simulations were quite similar to those observed in vitro.

In collagen, the percentage of cysts with multiple lumens was larger than in Matrigel. The percentage of ISMA-C CYSTS with multiple LUMENS was similar to that observed within collagen (Fig. 4A). In addition, the percentage of CYSTS with a single lumen surrounded by a single layer of CELLS (ideal cysts) was also quite similar to in vitro percentages (Fig. 4B). Delaying the formation of ISMA-C LUMENS caused the percentage of CYSTS with multiple LUMENS to increase sufficiently to match in vitro percentages without further parameter adjustments.

Cells within cysts in collagen are more likely to undergo apoptotic cell death than are counterparts in Matrigel [1]. The percentage of ISMA-C CYSTS in which DYING CELLS were observed (results not shown) did not significantly differ from earlier, published ISMA-M percentages and thus failed to match percentages observed in collagen. Future ISMA-C experiments are needed to explore the consequences of altering parameters that control CELL DEATH in order to obtain frequencies that better match in vitro, while retaining the already achieved attribute similarities.

In order to mimic the differences in cell number and onset of lumen formation observed within collagen cultures, three ISMA-M parameters were altered. The resulting ISMA-C not only mimicked quantitative measures for cell number and lumen initiation in collagen cultures, it also mimicked behaviors for cyst size,

lumen size, mean cell size, the ratio of cellular to cyst area, the percentage of cysts with single and multiple lumens, and the percentage of single lumen, single layer cysts.

5. DISCUSSION

In this report we demonstrate use of the Iterative Refinement Protocol. ISMA-M, an analogue developed previously to mimic quantitative measures of MDCK cell growth in Matrigel, was falsified by addition of measures of cystogenesis in collagen to its targeted attribute list. Parameters within the ISMA-M were modified to allow the resulting ISMA-C to survive falsification by most of the new data. Measures taken during ISMA-C simulations quantitatively mimicked corresponding measures within collagen. We believe it is informative to note which parameters had to be altered in order to produce ISMA-C, and which could be left unchanged. To mimic cell numbers, parameters controlling both the initial CELL number and CELL DIVISION rate were altered. With further refinement, we may learn that a higher rate of CELL death will produce later stage numbers that match the lower in vitro cell numbers. However, it is unlikely to be the case during the first four days of growth.

In order to mimic observed values for the onset of lumen formation, the ISMA-M value of *polarDelay* was increased. So doing was in agreement with the hypothesis proposed in [1]: the primary difference between cyst growth in Matrigel and collagen is due to delayed polarization (although our simulation also predicts a decrease in the rate of cellular division). We noted that values of CYST and LUMEN size generated during ISMA-C simulations were remarkably close to values observed within collagen, even though ISMA-C were not modified specifically to achieve that result. That evidence supports our hypothesis that ISMA-C mechanisms have in vitro counterparts.

ISMA-C simulations achieved other validation targets without further modification. We take this as additional evidence that ISMA-C operating principles do have in vitro MDCK counterparts. The percentages of CYSTS with single or multiple LUMENS were remarkably similar to in vitro percentages. Although ISMA-C CELLS POLARIZED and formed LUMENS later than in ISMA-M simulations, the number of ISMA-C CELLS within clusters when LUMENS first appeared was similar to that observed in ISMA-M simulations. It may be that the timing of the onset of lumen formation in vitro is primarily a function of cell number.

In order to implement these simulations using agent-oriented modeling techniques, it was necessary to reengineer and expand certain aspects of the modeling framework. Cellular Potts models are useful because they can simulate aspects of cell biology like cell size and shape change, but they are not fundamentally agent-oriented. By expanding the CompuCell3D framework to allow direct execution of the CELL agents, we changed the perspective used to create the underlying analogue. Using a perspective consistent with prior simulation efforts allowed concepts and techniques such as the Iterative Refinement Protocol to be fully utilized. It also allowed for the direct mapping from in vitro to in silico operating principles and from in silico observations to in vitro hypotheses. By executing individually and only requiring information about their immediate neighbors, CELLS maintain a degree of autonomy. Increasing the quasi-

autonomous nature of agents within simulations will encourage good modeling practices and make mapping from in silico to in vitro more natural and direct.

ISMA-C CELLS stabilize later, at roughly day 8, rather than at day 6 as observed in vitro. This difference indicates that either the ISMA-C mechanisms are a flawed representation of in vitro counterparts, or that a different value of LUMEN size is needed. Further, ISMA-C simulations do not mimic the increased rate of cell death observed in vitro. Additional iterative refinement is needed to discover an ISMA-C that does not exhibit those discrepancies. Modification of *stableRatio* is expected to cause CELLS to stabilize earlier. Changes in the parameters influencing CELL DEATH will increase the frequency of CELL DEATH, but that change will impact those features that currently validate. Further successful refinement will expand ISMA-C's phenotype so that it covers a larger portion of the MDCK culture system's in vitro phenotype. The current results support the hypothesis that in achieving cystogenesis, MDCK cells use the same basic operating principles when cultured in Matrigel or collagen.

6. IMPLEMENTATION TOOLS

The model was implemented using CompuCell3D v 3.2.1 (http://compucell3d.org/) and custom code. CompuCell3D is a grid-based cellular Potts model simulation framework. Simulation data was stored in a MySQL v 5.0 (http://www.mysql.com/) database and analyzed using R v 2.10.1 (http://www.r-project.org). Simulations were executed using a Dell Poweredge 1900 server with two 4-core 2.33 GHz 64 bit Intel Xeon processors, 8 GB of RAM and a 450 GB hard drive. The system software was Ubuntu 8.04 LTS (Linux kernel 2.6).

ACKNOWLEDGMENTS

We would like to thank members of the UCSF BioSystems group and the Mostov lab for helpful suggestions and discussion. We gratefully acknowledge research funding provided by the CDH Research Foundation.

REFERENCES

[1] Martín-Belmonte, F.; W. Yu; A.E. Rodríguez-Fraticelli; A.J. Ewald; Z. Werb; M.A. Alonso; K. Mostov. 2008. "Cell-Polarity Dynamics Controls the Mechanism of Lumen Formation in Epithelial Morphogenesis." *Curr Biol* 18:507-513.

[2] Engelberg, J.A.; A. Datta; K.E. Mostov; C.A. Hunt. 2011. "MDCK Cystogenesis Driven by Cell Stabilization Within Computational Analogues." *PLoS Comput Biol (submitted)*.

[3] Cickovski T.M.; C. Huang; R. Chaturvedi; T. Glimm; H.G. Hentschel, et al. 2005. "A Framework for Three-Dimensional Simulation of Morphogenesis." *IEEE/ACM Trans Comput Biol Bioinform* 2(4): 273-288.

[4] Kim, S.H.; J. Debnath; K. Mostov; S. Park; C.A. Hunt. 2009. "A Computational Approach to Resolve Cell Level Contributions to Early Glandular Epithelial Cancer Progression." *BMC Syst Biol* 3:122.

[5] Hunt, C.A.; G.E. Ropella; T.N. Lam; J. Tang; S.H. Kim; J.A. Engelberg; S. Sheikh-Bahaei. 2009. "At the Biological Modeling and Simulation Frontier." *Pharm Res* 26(11):2369-2400.

[6] Tang, J.; C.A. Hunt. 2010. "Identifying the Rules of Engagement Enabling Leukocyte Rolling, Activation, and Adhesion". *PLoS Comput Biol* 6(2):e1000681.

[7] Engelberg, J.A.; G.E.P. Ropella; C.A. Hunt. 2008. "Essential Operating Principles for Tumor Spheroid Growth." *BMC Sys Biol* 2:110.

[8] Glazier J.A.; A. Balter; N.J. Poplawski. 2007. "Magnetization to morphogenesis: A brief history of the Glazier-Graner-Hogeweg model. In: Rejniak KA, editor. Single-Cell-Based Models in Biology and Medicine. Basel: Birkhäuser. pp. 79-106.

[9] Graner F.; J.A. Glazier. 1992. "Simulation of Biological Cell Sorting Using a Two-Dimensional Extended Potts Model." *Phys Rev Lett* 69(13): 2013-2016.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Jess Engelberg

Author Signature

3/11/2011

Date