Lean-Agile city.

This place runs on folklore, intuition, and anecdotes.

If you want to know the truth about this town, stick with me. I’ll give you a tour you’ll never forget.

But if you don’t want your beliefs challenged with facts, you’d better beat it, kid. I don’t want to upset you.
My sidekick down there? That’s Larry Maccherone. He’s worked in this town his entire professional life.
I’m going to give you the tools to find the real-world numbers that can help you make the economic case to get the resources you need and get your people to commit to change. Really.
# The Seven Deadly Sins of Agile Measurement

1. **Manipulating Others**
2. **Unbalanced Metrics**
3. **Quantitative Idolatry**
4. **Overpriced Metrics**
5. **Lazy Metrics**
6. **Bad Analysis**
7. **Linear Forecasting**
Sin #1

Manipulating Others

Using metrics as a lever to drive someone else’s behavior
Heavenly Virtue #1

Self Improvement

Using metrics to reflect on your own performance
Sin #7

Linear Forecasting

Forecasting without discussing probability and risk
Heavenly Virtue #7

Probability Tools

Using the proper tools to predict the likelihood of results
Monte Carlo Simulation
CAUTION:
Correlation does not necessarily mean causation
CAUTION:
There are no best practices
Only good practices in context
The investigation continues with ...

Iteration length
Crowd wisdom or shared delusion?

<table>
<thead>
<tr>
<th>Iteration length</th>
<th>Teams using</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 week</td>
<td>6.2%</td>
</tr>
<tr>
<td>2 weeks</td>
<td>59.1%</td>
</tr>
<tr>
<td>3 weeks</td>
<td>23.4%</td>
</tr>
<tr>
<td>4 weeks</td>
<td>9.8%</td>
</tr>
<tr>
<td>5+ weeks</td>
<td>1.5%</td>
</tr>
</tbody>
</table>
SDPI current dimensions

- **Productivity (Throughput)**
- **Predictability (Stability of Throughput)**
- **Responsiveness (Time in Process)**
- **Quality (Defect Density)**
Future SDPI dimensions

- Customer/Stakeholder Satisfaction (Late 2014)
- Build-the-Right-Thing metric (2015)
- Employee Engagement/Satisfaction (Late 2014)
- Code Quality from Static Analysis (2015)
Raw metrics → Percentiles = Index
The investigation continues with ...

Iteration length
PERFORMANCE

Iteration Length relationship to Performance

<table>
<thead>
<tr>
<th>Iteration Length</th>
<th>Productivity</th>
<th>Predictability</th>
<th>Responsiveness</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 week</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Performance index total

©2014 Rally Software
@LMaccherone | LMaccherone@rallydev.com | #Agile2014
PRODUCTIVITY
Iteration Length relationship to Performance

Performance index

1 week  2 weeks  3 weeks  4 weeks  5+ weeks

Iteration Length

©2014 Rally Software
QUALITY
Iteration Length relationship to Performance

Performance index

1 week | 2 weeks | 3 weeks | 4 weeks | 5+ weeks

Iteration Length
RESPONSIVENESS

Iteration Length relationship to Performance

![Graph showing the relationship between iteration length and performance index. The x-axis represents iteration lengths of 1 week, 2 weeks, 3 weeks, 4 weeks, and 5+ weeks, while the y-axis represents the performance index ranging from 0 to 90. The graph indicates that shorter iteration lengths (1 week) tend to have higher performance indices compared to longer iteration lengths.]
PREDICTABILITY

Iteration Length relationship to Performance

[Graph showing performance index for different iteration lengths (1 week, 2 weeks, 3 weeks, 4 weeks, 5+ weeks).]
Iteration Length Transition vs Avg Responsiveness

Avg Responsiveness

Month_{-3}  Month_{-2}  Month_{-1}  Month_{0}  Month_{+1}  Month_{+2}  Month_{+3}

2 weeks→3 weeks (107)  3 weeks→2 weeks (86)
Iteration Length Transition vs Avg Performance Index

- 2 weeks→3 weeks (43)
- 3 weeks→2 weeks (32)
Facts Discovered:

- Teams using two-week iterations have the best balanced performance
- Longer iterations correlate with higher Quality
- Shorter iterations correlate with higher Productivity and Responsiveness
- However, some teams are acting like “tough guys” by pretending to operate at one-week iterations when they can’t back it up
The investigation continues with ...

Survey-based research
Where in the world?
Where in the Europe?
The investigation continues with...

Ratio of testers to developers
TESTERS PER DEVELOPER

Counts of measurements in each bucket
PRODUCTIVITY
Testers per Developer relationship to Performance

Performance index

Testers per Developer

0
< 0.3
0.3–0.6
0.6–1
QUALITY

Testers per Developer relationship to Performance

Performance Index

Testers per Developer

©2014 Rally Software
RESPONSIVENESS
Testers per Developer relationship to Performance

Performance Index

0  10  20  30  40  50  60  70  80  90

Testers per Developer

0  < 0.3  0.3-0.6  0.6-1

©2014 Rally Software
@LMaccherone | LMaccherone@rallydev.com | #Agile2014
PREDICTABILITY

Testers per Developer relationship to Performance

Performance index

Testers per Developer

©2014 Rally Software

@LMaccherone | LMaccherone@rallydev.com | #Agile2014

©2014 Rally Software
Facts Discovered:

- More testers lead to better Quality
- But they also generally lead to worse Productivity and Responsiveness
- Interestingly, teams that self-identify as having no testers have:
  - The best Productivity
  - Almost as good Quality
  - But much wider variation in Quality
The investigation continues with ...

Motive
REASON FOR SUCCESS
Counts of measurements in each bucket

- Exceptional Skill/Talent/Experience (n = 112)
- Disciplined Process (n = 51)
- Great Teamwork (n = 422)
- Great Tooling (n = 4)
- Great Technology (n = 24)
- Excellent Market Fit (n = 27)

Performance Index total:
- Exceptional Skill/Talent/Experience
- Disciplined Process
- Great Teamwork
- Great Tooling
- Great Technology
- Excellent Market Fit

Legend:
- Green: Productivity
- Blue: Predictability
- Purple: Responsiveness
- Orange: Quality
Evidence Found:

- Motive has a small but statistically significant impact on performance
- Extrinsic motivation does not have a negative impact on performance
- Executive support is critical for success with Agile
- Teamwork is not the dominant factor; talent, skills, and experience are
- Those motivated by quality perform best
The investigation continues with ...

Retrospectives
PERFORMANCE

Has Sprint Retrospective relationship to performance

<table>
<thead>
<tr>
<th></th>
<th>Productivity</th>
<th>Predictability</th>
<th>Responsiveness</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disagree (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly Agree (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Performance index total
HAS SPRINT RETROSPECTIVE
Counts of measurements in each bucket

- Strongly Disagree (1) (n = 15)
- Disagree (2) (n = 42)
- Neutral (3) (n = 57)
- Agree (4) (n = 204)
- Strongly Agree (5) (n = 437)
PERFORMANCE
Has Sprint Retrospective relationship to performance
PERFORMANCE
Has Sprint Retrospective relationship to performance

Performance index score

Strongly Disagree (1)  Disagree (2)  Neutral (3)  Agree (4)  Strongly Agree (5)

Productivity  Predictability  Responsiveness  Quality
PERFORMANCE

Has Sprint Retrospective relationship to performance

Performance index score

Strongly Disagree (1)  Disagree (2)  Neutral (3)  Agree (4)  Strongly Agree (5)

Productivity  Predictability  Responsiveness  Quality
PERFORMANCE

Has Sprint Retrospective relationship to performance

Performance index score

Strongly Disagree (1)  Disagree (2)  Neutral (3)  Agree (4)  Strongly Agree (5)

- Productivity
- Predictability
- Responsiveness
- Quality
The investigation continues with ... 

Co-location
Evidence Found:

- Teams distributed within the same time zone have up to 25% better productivity.
- Is distraction a problem?
The investigation continues with ... Geography
PERFORMANCE
Geographic Location relationship to Performance

Performance index total

<table>
<thead>
<tr>
<th>Geographic Location</th>
<th>Productivity</th>
<th>Predictability</th>
<th>Responsiveness</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia + NZ</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Americas</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Europe</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Israel</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pacific</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>India</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Israel-based teams
- Find more defects overall
- But find fewer in production
- Theory: May correlate with high use of static analysis tools

India-based teams
- Find more defects overall
- Released and unreleased
- Theory: May correlate with high use of static analysis tools
- Theory: Could be recording bias
Facts Discovered:

- Differences are slight but statistically significant.
- Australia has the best overall performance.
- India the worst. However, there could be a reporting bias for defects.
- Israel seems to catch the most defects before production. Heavy use of static analysis?
One year earlier ...

rallydev.com/agilemetrics
Teams with low WiP have up to:
- 4x better Quality
- 2x faster Time to market
- But 34% worse productivity

Stable teams result in up to:
- 60% better Productivity
- 40% better Predictability

Dedicated teams: Teams made up of people who only work on that one team have double the Productivity

Smaller teams have better Productivity
Larger teams have better Quality
What’s next?

- **Demo of Rally Insights**
  - Implements SDPI

- **Roadmap**
  - Self-assessment and tracking (surveys)
    - Two more dimensions
      - Customer/Stakeholder Satisfaction
      - Employee Engagement/Satisfaction
    - Probe your environment with customized surveys (maturity, practices compliance, etc.)
  - Recommendation Engine
    - What are the top five things we should improve next?

- Visit the Rally booth to get questions answered

- Get a sticker for Rally Survival Handbook to win a Google Glass
Over 55 variables under study
A fact without a theory is like a ship without a sail, is like a boat without a rudder, is like a kite without a tail. A fact without a figure is a tragic final act. But one thing worse in this universe is a theory without a fact.

~ George Schultz
Replace Folklore with Facts

Swap Anecdotes with Evidence

Upgrade Intuition to Insights
Additional slides
The investigation continues with ...

SDPI dimensions
Productivity = Throughput

Throughput is simply the count of User Stories completed in a given time period.

Productivity (by default) is the percentile scoring of the raw Throughput metric for User Stories normalized by team size.
Predictability = Stability of Throughput

Predictability measures how consistent you are at producing the same amount of work each month as measured by the Coefficient of Variation (CoV) of Throughput.

Predictability (by default) is the percentile scoring of the raw CoV of Throughput.
Responsiveness = Time in Process

TiP shows how long it takes to get one work item through your system. It's the work days that a User Story spends in development and testing. Similar to lead time or cycle time.

Responsiveness (by default) is the percentile scoring of the raw Time In Process (TiP) metric for User Stories.
Quality = Defect Density

Defect Density is a representation of the number of defects found in your code. It’s the count of defects found in a given time period, normalized by team size.

Quality (by default) is the percentile scoring of the raw defect density metrics for both defects found in test as well as those found in production.
The investigation continues with...

Team time together
PRODUCTIVITY
Team Time Together relationship to Performance

Performance index

Team Time Together

- <= 1 Month
- > 1 Month <= 3 Months
- > 3 Months <= 6 Months
- > 6 Months <= 1 Year
- > 1 Year <= 3 Years
- > 3 Years
PREDICTABILITY

Team Time Together relationship to Performance

Performance index

Team Time Together

<= 1 Month
> 1 Month <= 3 Months
> 3 Months <= 6 Months
> 6 Months <= 1 Year
> 1 Year <= 3 Years
> 3 Years
The investigation continues with ... Location in US and Europe
**PRODUCTIVITY**

US Geographic Location relationship to Performance

![Bar chart showing productivity across different US geographic locations](chart.png)
PRODUCTIVITY

Europe Geographic Location relationship to Performance

Performance index

Europe Geographic Location

Denmark  Belgium  England  France  Germany  Switzerland  Netherlands  Italy  Sweden  Norway  Bulgaria
The investigation continues with ...

Controlling WiP
Most obvious finding: Little’s Law

RESPONSIVENESS
Work in Process (WIP) per Person relationship to Performance
QUALITY

Work in Process (WIP) per Person relationship to Performance

Most dramatic finding

©2014 Rally Software
@LMaccherone | LMaccherone@rallydev.com | #Agile2014
©2014 Rally Software
PREDICTABILITY
Work in Process (WIP) per Person relationship to Performance

Performance Index

Work in Process (WIP) per Person

0 - 1
1 - 2
2 - 3
3 - 5
5 - 7
> 7
PRODUCTIVITY

Work in Process (WIP) per Person relationship to Performance
Facts Discovered:
Teams that most aggressively control WiP:

- Have $\frac{1}{2}$ the Time in Process (TiP)
- Have $\frac{1}{4}$ as many defects
- But have 34% lower productivity
Recommendations:

- If your WiP is high, reduce it.
- If your WiP is already low, consider your economic drivers:
  - If productivity drives your bottom line, don’t push WiP too low.
  - If time to market or quality drives your bottom line, push WiP as low as it will go.
The investigation continues with ...

Estimating process
# Estimating process

<table>
<thead>
<tr>
<th>Process Type</th>
<th>Teams Using</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Estimates</td>
<td>3%</td>
</tr>
<tr>
<td>Full Scrum</td>
<td>79%</td>
</tr>
<tr>
<td>Lightweight Scrum</td>
<td>10%</td>
</tr>
<tr>
<td>Hourly-Oriented</td>
<td>8%</td>
</tr>
</tbody>
</table>
PERFORMANCE
Process Type relationship to Performance

Performance index total

<table>
<thead>
<tr>
<th>Process Type</th>
<th>Productivity</th>
<th>Predictability</th>
<th>Responsiveness</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lightweight Scrum</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Full Scrum</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>No Estimates</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Hourly</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>
QUALITY
Process Type relationship to Performance

![Chart showing performance index for different process types: Lightweight Scrum, Full Scrum, No Estimates, and Hourly.](chart.png)
Released Defect Density (Average)

Lower is better

Defects per 1000 Person-days

- Lightweight Scrum
- Full Scrum
- No Estimates
- Hourly
PRODUCTIVITY
Process Type relationship to Performance

- Lightweight Scrum
- Full Scrum
- No Estimates
- Hourly

Performance index
RESPONSIVENESS
Process Type relationship to Performance

Performance index

Lightweight Scrum
Full Scrum
No Estimates
Hourly

Process Type
Facts Discovered:

- Teams doing Full Scrum have 250% better Quality than teams doing no estimating.
- Lightweight Scrum performs better overall, with better Productivity, Predictability, and Responsiveness.
Recommendations:

- Experienced teams may get best results from Lightweight Scrum
- If new to Agile or focused strongly on Quality, choose Full Scrum
The investigation continues with ...

Team stability & Dedication to one team
PERCENT DEDICATED WORK
Counts of measurements in each bucket

- <= 50: 17,837
- 50 - 70: 14,837
- 70 - 85: 15,091
- 85 - 95: 16,414
- > 95: 81,472
PRODUCTIVITY

Percent Dedicated Work relationship to Performance

Performance Index

Percent Dedicated Work

<= 50
50 - 70
70 - 85
85 - 95
> 95
QUALITY

Percent Dedicated Work relationship to Performance

![Graph showing the relationship between percent dedicated work and performance index. The graph has four bars representing different ranges of percent dedicated work: <= 50, 50 - 70, 70 - 85, 85 - 95, and > 95. The bars indicate the performance index with error bars showing variability.](image-url)
PERCENT DEDICATED
Counts of measurements in each bucket
TEAM STABILITY
Counts of measurements in each bucket

Percent Dedicated

Count

©2014 Rally Software
Another Fact Discovered:
One out of four team members changes every three months!
PRODUCTIVITY

Team Stability relationship to Performance

Performance Index

Team Stability

0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
PREDICTABILITY

Team Stability relationship to Performance

Performance index

Team Stability

0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
QUALITY
Team Stability relationship to Performance

![Graph showing the relationship between team stability and performance index. The graph indicates that higher team stability is associated with higher performance index.](chart.png)
Facts Discovered:
Stable teams result in up to:
● 60% better Productivity
● 40% better Predictability

Another Fact Discovered:
One out of four team members changes every three months!
Recommendations:

- Dedicate people to a single team
- Keep teams intact and stable
The investigation continues with ...

Team size
Balance your team’s Performance

Agile recommends that the ideal team size is 7± 2. How ideal is that when we actually look at the data?
PERFORMANCE
Team Size relationship to performance

- Productivity
- Predictability
- Responsiveness
- Quality

Performance Index Total

- < 3
- 3 - 5
- 5 - 9
- 9 - 15
- 15 - 20
- > 20
PERFORMANCE

Team Size relationship to performance

![Performance Chart]

- **Productivity**
- **Predictability**
- **Responsiveness**
- **Quality**

Performance index score:
- **< 3**
- **3 - 5**
- **5 - 9**
- **9 - 15**
- **15 - 20**
- **> 20**
PERFORMANCE

Team Size relationship to performance

![Performance Chart]

- **< 3**: Productivity, Predictability
- **3 - 5**: Productivity, Predictability
- **5 - 9**: Productivity, Predictability
- **9 - 15**: Productivity, Predictability
- **15 - 20**: Productivity, Predictability
- **> 20**: Productivity, Predictability

©2014 Rally Software

@RallySoftware | @LMaccherone | LMaccherone@rallydev.com | #Agile2014 ©2014 Rally Software
PERFORMANCE

Team Size relationship to performance

<table>
<thead>
<tr>
<th>Team Size</th>
<th>Performance Index Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 3</td>
<td>75</td>
</tr>
<tr>
<td>3 - 5</td>
<td>50</td>
</tr>
<tr>
<td>5 - 9</td>
<td>50</td>
</tr>
<tr>
<td>9 - 15</td>
<td>50</td>
</tr>
<tr>
<td>15 - 20</td>
<td>50</td>
</tr>
<tr>
<td>&gt; 20</td>
<td>25</td>
</tr>
</tbody>
</table>

Legend:
- Productivity
- Predictability
- Responsiveness
- Quality
Facts Discovered:
Small teams (of 1-3) people have:
- 17% lower Quality
- But 17% more Productivity than teams of the recommended size.
Recommendations:

- Set up team size of 7±2 people for the most balanced performance
- If you are doing well with larger teams, there’s no evidence that you need to change