Predicting the Success of New Cable Series from Their Pilot Episode Scripts: An Empirical Approach

Starling D. Hunter[1], Yelitza Prada Breen[2]

[1]Carnegie Mellon University Qatar, Qatar

[2]UCLA Extension Film Program, USA

Correspondence: Starling D. Hunter, Carnegie Mellon University Qatar, Qatar. E-mail: starling@mit.edu

Abstract

In this study we develop an empirical model to explain failure rates of new television series. Specifically, we test the ability of three factors to predict the success of new dramatic series appearing on 31 cable networks over the last 10 years. Those factors are the originality of the story, the track record of its creator(s), and the cognitive complexity of its pilot episode script—all of which are known well in advance of a network's decision to greenlight a new series. As predicted, we find that all three variables—both individually and in combination—strongly predict the success rate of new dramatic series in their first two seasons.

Keywords: television, network analysis, pilot episodes, ratings, audience, dramatic series, content analysis, forecasting

1. Introduction

For the last few decades, perhaps even longer, first-season failure rates of new televisions series have consistently topped 60% while the failure rates within the first two seasons have often reached 80% (Stelter, 2012). For some networks and production companies, the high failures rates are the cost of doing business—and at $2-5 million dollars per greenlit pilot, it's an exorbitant cost (Nathanson, 2013). Somewhat surprisingly, academic researchers have shown very little interest in identifying the determinants of success and failure of new television series (Bielby & Bielby, 1994). Rather, the large majority of effort has been devoted to a different but somewhat related question—that of explaining variation in the *ratings* of television shows (Danaher & Dagger, 2012). And within that research stream, new shows have not been deemed a population worthy of exclusive focus. Recently this changed. Specifically, in three studies (Hunter, Smith, & Chinta, 2016; Hunter, Chinta, Smith, et. al, 2016; Hunter & Breen, 2017), three content analytic and textual factors extracted from pilot episode scripts were shown to predict the audience size and season length of the subsequent series, all of which appeared on the four major US television networks over the last decade. But while those studies comprise an important step in understanding the causes of success and failure of new televisions series, the generalizability of the results is somewhat limited. That's because the studies took no account of cable and streaming outlets—the fastest growing source of new television series (Littleton, 2015), as well as some of the most ground-breaking programming (Ryan, 2016). The express intent of this study is to address this gap in our understanding. Specifically, it is to determine whether the same three aforementioned factors predict commercial performance in new dramatic television series appearing on over 30 cable and streaming outlets over the last decade.

2. Literature Review and Hypotheses

Failure rates of new television series in their first season typically top two thirds, with around 80% being canceled by the end of their second season (Stelter, 2012; Nathanson, 2013; Porter, 2016). According to Stelter (2012), rarely, if ever, do industry insiders "question whether this is appropriate" Rather, it appears that high failure rates and the attendant "lavish spending" are accepted as the price of doing business or of "staying relevant" (Nathanson, 2013).

With one notable exception (Bielby & Bielby, 1994), academic research—at least that of the empirical variety—has ignored this question of high failure rates and associated development spending in favor of a different but related question—explaining variance in ratings of all shows being broadcast and not the subset of which are new on the air. Among the variables that this literature has examined and found predictive are time of day and season that the program is broadcast, the program's genre (sports, drama, comedy, etc.), the total number of programs broadcast during the study

period, the program's duration and whether it is a re-run, and the ratings of the program appearing in the time-slot immediately before (lead-in) and after (lead-out) (Danaher & Dagger, 2012). Notably, with the exception of genre, these are all post-production factors, i.e. ones known well *after* a decision was made to green-light the series.

The recent work of Hunter and colleagues (Hunter, Smith, & Chinta, 2016; Hunter, Chinta, Smith, et al, 2016; Hunter & Breen, 2017) has a focused directly on the question of determinants of the performance of new shows. Those studies draw on related work in film studies and cultural economics that has explained box office performance using only variables known to production executives during the pre-production stages (Eliashberg, Hui, & Zhang, 2007, 2014; Hunter, Smith & Singh, 2016), i.e. before a decision has been made to greenlight a film project. At this early stage, the most important factors are those derived from the script that forms the basis for the film or the new series. Three such factors have been shown in the work of Hunter and colleagues to strongly predict the performance of new dramatic series appearing on network television over the last decade. They are the originality of the series' premise, the track record of the series' creators, and the cognitive complexity of the pilot episode script.

Concerning the former, they found that new dramatic series with original concepts, i.e. ones not derived from a film, adapted from novel, or based on comic book series, etc. had significantly higher viewership in the first five episodes of their first season (Hunter, Smith, & Chinta, 2016), as well as for the first season as a whole (Hunter, Chinta, & Smith, et al, 2016). They also reported that new series with original concepts had a greater number of episodes in their first season and were more likely to have a full first season compared to new series with unoriginal concepts (Hunter & Breen, 2017). Accordingly, our first hypothesis is that

H1: All else equal, new cable series with *original premises* will outperform new cable series with unoriginal or adapted premises.

As for the second factor, in the same three studies mentioned above, it was reported that new series from creators with a proven track record of success outperformed new series whose creators lacked such a track record. These findings parallel those of Bielby & Bileby (1994), wherein it was reported that the "… single best predictor of whether a pilot is selected for prime-time schedule is whether or not it has been described as originating with an established producer of a prior successful series" (p. 1305). They did not find, however, that the impact of these claims predicted "the series' subsequent *commercial* success" (ibid, p. 1300, emphasis added). More specifically, they did not find that claims about "writer-producers" explained variation in either the series average ratings for the first season or whether the series "was selected to return for a second season in 1992-3" (ibid, p. 1308). Thus, while the conclusions of prior research are somewhat mixed, our second hypothesis is that

H2: All else equal, new cable series whose creators have a *track record* of success will outperform new cable series from creators without a track record.

As for the third factor, Hunter and colleagues have reported a positive relationship between the cognitive complexity of the pilot episode script and the performance of the subsequent series. Specifically, they found higher cognitive complexity to be associated with higher viewership in the first five episodes of new dramatic series (Hunter, Smith, Chinta, 2016), as well as the entire first season (Hunter, Chinta, Smith, et al, 2016). They also reported that new series with more cognitively complex scripts had a greater number of episodes in their first season and were more likely to enjoy a "full" first season compared to new series whose pilot scripts did not have high cognitive complexity (Hunter & Breen, 2017). Here "full" was taken to mean 20 or more episodes. Thus, our third hypothesis is that

H3: All else equal, new cable series whose pilot episode scripts have high *cognitive complexity* will outperform new cable series whose pilot episode scripts that do not have high cognitive complexity.

3. Data and Methods

We used a variety of online sources such as *Wikipedia*, *TV Guide*, *TV.com*, and *TV Series Finale*, to identify original dramatic series that debuted on any of several US-based, cable and/or streaming networks between September 1, 2007 and March 31st, 2017. After eliminating from further consideration those series (1) that were anthologies (2) whose pilot series scripts could not be located or were not machine readable (3) that were spun-off from existing series via "back-door" pilots or (4) were not US-produced or co-produced, some 198 pilots appearing on 31 different networks remained in our sample. The vast majority of pilot episode scripts for these series were obtained free-of-charge from the "TV Writing" website (https://sites.google.com/site/tvwriting/) with the remainder purchased from Scriptfly (https://www.scriptfly.com/).

3.1 Dependent Variable

Because viewership and audience size data is not available for cable networks in the same way as it is for network series, we were unable to directly replicate the performance measures used by Hunter, Smith, and Chinta (2016) or Hunter, Chinta, Smith, et al (2016). And because season lengths are routinely shorter on cable networks and because the

standard "full" season is similarly shorter, we chose not to measure performance in the manner of Hunter & Breen (2017) which was the number of episodes in the first season and the likelihood of a new series getting a "full" first season. Instead, we opted for two different but related measures of performance, none of which had been previously tested. The first, named CANCEL1, was a categorical variable coded "1" if the show was canceled after its first season and coded "0" otherwise. The second variable, CANCEL2, was also a categorical variable coded "1" if the show was canceled during or at the end of its second season, coded "0" otherwise. We were able to obtain information needed to calculate CANCEL1 for 198 series and information to calculate CANCEL2 for 172 of them. As shown in Table 1, 54 of 198 series (27%) of series were canceled at the conclusion of their first season while 47% (81 of 172) were canceled within the first two seasons. These numbers are dramatically lower than in network television where 2/3rd are routinely canceled in year 1 and 4/5th by the end of the second.

Table 1. Descriptive Statistics

Variable	N	Mean	Std. Dev.	Min	Max
DEPENDENT VARIABLES					
CANCEL1	198	0.27	0.45	0	1
CANCEL2	172	0.47	0.50	0	1
INDEPENDENT VARIABLES					
ORIGINAL	198	0.54	0.50	0	1
TRACK	198	0.26	0.44	0	1
COMPLEXITY	198	0.22	0.42	0	1
CONTROL VARIABLES					
Year of Release (YEAR)	198	2013.1	2.39	2007	2017
Genre = Crime (CRIME)	198	0.28	0.45	0	1
Log of # of Episodes in 1st Season (LOGEPS)	198	1.06	0.14	0.30	1.48
Series in Sample by Network > 10 (NETGT9)	198	0.58	0.49	0	1

3.2 Independent Variables

3.2.1 Story Originality

We determined the originality of each series by consulting its *Wikipedia* and/or *International Movie Database* (IMDb) page wherein we looked for information indicating if the story was derived from prior source material. Series were considered to be original if and only if they were (1) not an adaptation of a *novel* and (2) not a remake of a *foreign television series* and (3) not a *spin-off* of a pre-existing series, a pre-quel or sequel, or franchise extension, and (4) not a *revival*, re-make, or re-imagining of a previously broadcasted show and (5) not an *adaptation for television* of a feature film or documentary and (6) not based on a *comic series* or a *fairy tale*. Additionally, we did not treat as original those series whose characters and basic storylines were previously described in either contemporary literature, in religious scripture, or classical mythology. To reflect this information, we created the categorical variable ORIGINAL that was coded "1" for series that met the aforementioned originality conditions and coded "0" otherwise. As noted in Table 2, 54% of series were classified as original. This quantity is 11% lower than the rate of original series reported by Hunter & Breen (2017) in their examination of a decade's worth of new series appearing on the four major broadcast networks.

3.2.2 Creator Track Record

In order to determine the track record of a series' creators we relied exclusively on information contained in the *International Movie Database* (IMDb). In our search of this database we identified a total of 295 creators associated with the 198 series in our dataset, an average of 1.38 creators per series. On occasion, there were creators listed for series who did not have a "written by" credit for the pilot episode but instead a "created by" or "story" credit. Those creators were excluded from our analysis and this brought the total number of down to 273. Next, for each series we first determined whether the creator(s) with "written by" credits had previously earned such a credit for any other series' pilot episode. We then checked the corresponding "Writer" page for each creator to determine which of those series, if any, had been renewed for a second season. Those renewed series could have appeared on either of the four major networks or any of dozens of US-based cable networks. We then created a variable named "TRACK" which was coded "1" if the series' creator(s) had written the pilot for at least one other show that had been renewed for a second or subsequent seasons. TRACK was coded "0" if this condition was not met. As indicated in Table 2, 26% of the series' had at least one creator with such a track record of success.

Notably, sixteen of the creators were associated with more than one series in our sample. These were Greg Berlanti (*Arrow*, *The Flash*, and *The Tomorrow People*), the writing team of Miles Milar and Alfred Gough (*Into the Badlands* & *The Shannara Chronicles*), Andrew Kreisberg (*Arrow* and *The Flash*), Craig Silverstein (*Turn* and *Nikita*), David Kelley (*Goliath* and *Monday Mornings*), Gideon Raff (*Tyrant* and *Dig*), Howard Gordon (*Homeland* and *Legends*), Jeff Eastin (*White Collar* and *Graceland*), Jennie Snyder Urman (*Jane the Virgin* and *Emily Owens, M.D.*), Julie Plec (*The Vampire Diaries*, *Containment*, and *The Tomorrow People*), Kurt Sutter (*Sons of Anarchy* and *The Bastard Executioner*), Marti

Noxon (*Unreal* and *The Girlfiends Guide to Divorce*), Noah Hawley (*Fargo* and *Legion*), Rob Thomas (*90210* and *iZombie*) , Ronald D. Moore (*Outlander* and *Caprica*), and Vince Gilligan (*Breaking Bad* and *Better Call Saul*).

3.2.3 Cognitive Complexity

Cognitive complexity is a concept in the psychology literature that has several definitions, two of which are 'the number of independent constructs a person uses in perceiving and interpreting the environment' (Tinsley, Kass, Moreland, & Harrison, 1983: 94) and 'the number of independent dimensions of concepts that an individual brings to bear in describing a particular domain of phenomena' (Scott, 1962: 405). Following prior research (Hunter, Smith, & Singh, 2016), we measured the cognitive complexity of a new series by the number of links contained in the main component of the concept network constructed from the screenplay of its pilot episode. In short, the greater the number of links, the greater the cognitive complexity. The concept network was constructed in several steps, the first of which was to identify all of the multi-morphemic compounds in the text of each series' pilot episode script. According to Hunter (2014), MMCs include, but are not necessarily limited to, *closed compounds* (milkshake, handheld), *hyphenated-compounds* (high-powered, right-handed), acronyms and abbreviations (NBA, FBI, NATO), *portmanteaus* or blend words (motel, medivac, helipad), *clipped words* (pub, cell, internet, taxi), and *pseudo-compounds* (overcooked, understand, under-estimate).

The second step was to trace each word within the MMC back to its most remote etymological root, most of which were "Indo-European" (Watkins, 2011). Next every MMC in a pilot script was represented as one or more links between two or more concepts (etymological roots). Finally, the entire set of concepts and links was rendered as a network graph using the UCINet (Borgatti, Everett, & Freeman, 2002) and its embedded Netdraw module. In Figure 1, below, there is presented a concept map for *Kingdom* (2014), a drama built around the owner of a mixed martial arts (MMA) gym in Venice, California and which is now in its fourth season on the Audience Network. The concept map has 86 nodes and 92 links in its network, well within the top 25% of the sample. Each link represents a multi-morphemic compound word (MMC) that appears in the script. Several such MMCs were contained in the pilot script, many of which would be expected for a drama about mixed martial arts. These included, but were not limited to, the following closed compounds (headbutt, lightweight, tapout, headgear, overhand, and workout) and hyphenated compounds (pent-up, cut-man, sauna-suit, man-handle, and hand-pads) and the acronym UFC (United Fighting Championship). Across the entire sample, the average number of links in such concept maps was 52 with a standard deviation of 47. The max was 225 links in the network of the spy-drama *Covert Affairs* (2010) which ran for five seasons on the USA Network. Among the smallest concept networks were those for the musical drama *Sex Drugs and Rock & Roll*, and three period pieces—the Roman-era *Spartacus*, the revolutionary war–era *Turn*, and the Civil War-era *Hell on Wheels*. Our measure of the cognitive complexity of each pilot episode script was a categorical variable, COMPLEXITY that was coded "1" if the number of links was greater than 87, which was the top quartile in the sample, and coded "0" otherwise.

3.3 Control Variables

3.3.1 The "Crime" Genre

The first of our four controls was a categorical or "dummy" variable named CRIME. It was coded "1" if IMDb classified the series as belonging to the "crime" genre and coded "0" otherwise. IMDb defines the crime genre as one in which "the protagonists or antagonists are criminals" and where we observe "numerous consecutive and interrelated scenes of characters participating, aiding, abetting, and/or planning criminal behavior or experiences usually for an illicit goal" (IMDb, 2017). As shown in the descriptive statistics, 28% of the 198 new dramatic series belonged to this genre.

3.3.2 Calendar Year

To determine whether there has been any longer term trend in our dependent measures, we included as a control variable the calendar year in which the show was first aired. As shown in the table of descriptive statistics, the average year in the sample was 2013.1 with a low of 2007 and a high of 2017.
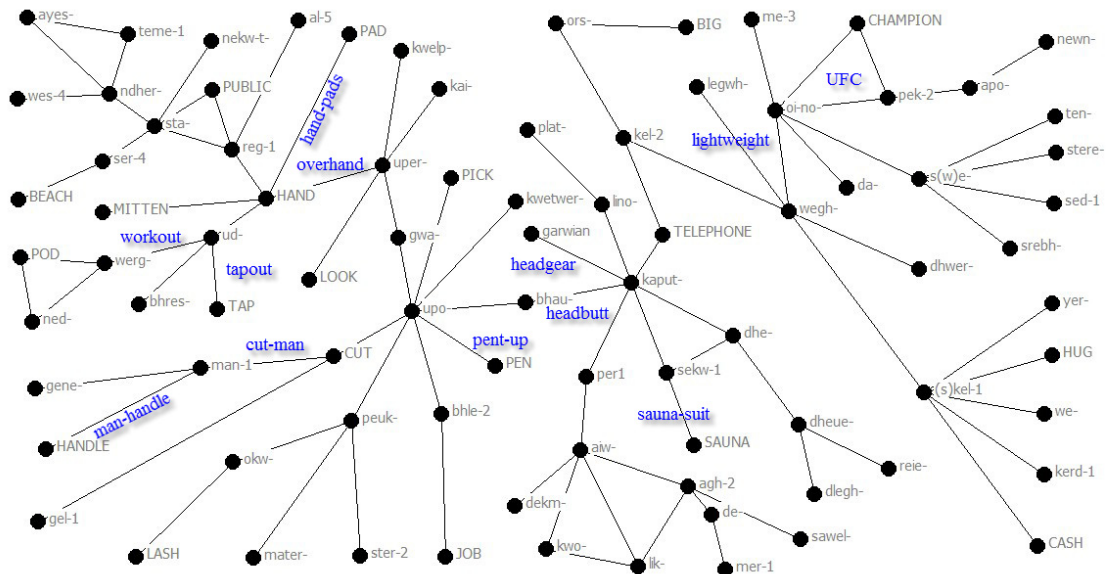
Figure 1. Concept Map for *Kingdom* (2014)

3.3.3 Episodes in Series' First Year

The series in our sample varied in terms of the number of episodes they had in their first season on the air. The high was thirty episodes for ABC Family's *Switched at Birth* (2011) while the low was just two episodes for the CW's fashion-centric drama *TBL: The Beautiful Life* which was pulled for exceptionally poor ratings. Also coming in at the lower end of the scale were three series from AMC—*The Walking* Dead (2010), *Fear the Walking Dead* (2015), and *Into the Badlands* (2015), another three from Sundance TV—*Hap & Leonard* (2016), *Rectify* (2013) and *The Red Road* (2014)—plus TNT's *Mob City* (2013).To control for possibility that the number of episodes in the first season had an effect on whether the show got renewed for subsequent seasons, we created a variable LOGEPS whose value was the log of the number of episodes appearing on the air in the first season.

3.3.4 Network Size

As noted in Table 1, series in our sample appeared on 31 different cable networks or streaming services. There was marked variation in the number of new dramatic series that these network aired over the nearly ten-year sample period. Almost a third of the networks had only one or two shows in the sample. These included the Oprah Winfrey Network, E!, Black Entertainment Television, WeTV, Bravo, TBS, Audience, PlayStation Network, The History Channel, and The Cartoon Network. At the other end of the scale were The CW with 28 series, ABC Family (now Freeform) with 17, the USA Network with 18, TNT with 16, and the American Movie Channel (AMC) with 13 series. The median number was four (4) series. To capture the effect of these differences, we created a categorical variable NETGT9 which was coded "1" if the network on which a series appeared had more than nine observations in our sample and coded "0" otherwise. As indicated in the descriptive statistics, the average of this variable was 0.58 while 0.49 was the value of the standard deviation.

4. Results

Recall that our three hypotheses were that the originality of a new series' premise, the possession of a track record success by the series creator(s), and the cognitive complexity of the pilot episode script would be strongly and positively associated with the performance of the series as a whole. To test these hypotheses, we created two measures of series performance: (1) whether or not the series was cancelled in its first season (CANCEL1) and (2) whether the series was cancelled within its first two seasons (CANCEL2). Table 2 below presents the results of four logistic regression models developed to test these three hypotheses. In each case the first model contains only the four control variables—YEAR, LOGEPS, CRIME and NETGT9. The dependent variable was CANCEL1. Model 1 in Table 3 indicates that only one of the four controls was statistically significant, NETGT9 (β= 1.02, z = 2.78, p < 0.01). The sign on this variable's regression coefficient indicates that series airing on the larger networks, e.g. The CW, USA Network, ABC Family, etc. were significantly *more* likely to be cancelled in their first season. In Model 2 the three independent variables were added—ORIGINAL, TRACK, and COMPLEXITY. Notably, the coefficient for ORIGINAL is positive but only marginally significant statistically (β= 0.50, z = 1.39, p< 0.10, 1-tailed), thereby indicating that original series were slightly *more* likely to be canceled after season 1. This is the *opposite* of our expectation which was based on

Hunter, Smith & Chinta (2016) who found that originality positively associated with higher viewership numbers and presumably a *lower* chance of first season cancelation.

Table 2. Logistic Regression where the Dependent Variable is a Categorical Variable Coded "1" if a Series was Cancelled after One Season and Coded "0" Otherwise

	Model 1	Model 2	Model 3	Model 4
YEAR	0.01	0.06	0.05	0.05
	(0.14)	(0.78)	(0.73)	(0.71)
LOGEPS	-0.52	-0.72	-0.36	-0.32
	(-0.44)	(-0.59)	(-0.30)	(-0.18)
CRIME	-0.08	0.05	0.02	0.01
	(-0.23)	(0.12)	(0.04)	(0.04)
NETGT9	1.02**	1.26***	1.21***	1.15***
	(2.78)	(3.23)	(3.13)	(3.02)
ORIGINAL (H1)		0.50#		
		(1.39)		
TRACK (H2)		-0.70#		
		(-1.62)		
COMPLEXITY (H3)		-1.84***		
		(-3.21)		
SUM			-0.82***	
			(-3.50)	
SUM1				-0.61*
				(-1.64)
SUM23				-1.72***
				(-3.30)
N	198	198	198	198
Chi-sq	8.4#	27.2***	22.1***	21.4***
Model df	4	7	5	6
Log Like	-111.8	-102.0	-105.0	-105.3
Pseudo-R2	3.6%	11.8%	9.5%	9.2%

Legend: # p< 0.10, 1-tailed, * p < 0.05, ** p< 0.01, *** p< 0.001 all 1-tailed

The coefficient for TRACK, the measure of the creative team track record, was, as expected, negative but only marginally significant (β= -0.70, z = 1.62, p < 0.10). Thus, series whose creator(s) had a track record of success were at lower risk of first season cancelation. Finally, the coefficient for COMPLEXITY was, as expected, negative and highly significant (β= -1.84, z = -3.21, p < 0.001). This indicates that pilot episode scripts that had high cognitive complexity were at much lower risk of being cancelled in or after their first season.

In Model 3 we introduced a new variable, SUM, which was the arithmetic sum of COMPLEXITY, TRACK, and the reverse of the value of ORIGINAL. In the case of the latter variable, wherever ORIGINAL assumed a value "0" we added a value of "1" to SUM, and vice versa. The resulting variable, SUM, captures whether there was an additive effect of these three independent measures. As shown in Model 3, the coefficient on SUM is, as expected, negative and highly significant (β= -0.82, z= -3.50, p < 0.001). This indicates that, all else equal, the presence of more of the three factors was associated with a lower likelihood of first season cancellation. Model 4 takes this analysis one step further and subdivides SUM into two categorical variables—SUM1 and SUM23. The former is coded "1" when SUM was equal to one and "0" otherwise while the latter is coded "1" when SUM was greater than 1 and "0" otherwise. As expected, the coefficients for both models are negative and statistically significant. That the magnitude of the coefficient for SUM23 is larger and more significant suggests that series having two or three factors present were much less likely than those with one or no factors present to be cancelled in or right after the first season.

Table 3, below, summarizes the raw cancellation rates across categories of independent and dependent measures. Notably, only 9% of series (1 in 11) whose pilot episodes were deemed cognitively complex experienced first season cancellation. However, about 31% of series whose pilots were not cognitively complex experienced first season cancelation. That's almost a 3.5-fold difference. Also of note is that 38% of series were cancelled in season 1 when none of the three factors was present while only 12% of those were cancelled when more than one was present. That's also more than a 3-fold difference. An even more dramatic decrease occurs for the 22 series where both SUM23 and COMPLEXITY were equal to 1: only one was cancelled in the first season, a rate of just 4.5% —an 9-fold difference compared to series with none of the three factors present. The one series that was canceled was an Amazon Studios period-piece entitled *Good Girls Revolt* (2016) that followed a "group of young female researchers" working at a news magazine modeled on Newsweek in the late 1960's.

Table 3. Independent Variables and Cancelation Rates

	% Cancelled in First Season	% Cancelled in First two Seasons
ORIGINAL = 0	23%	38%
ORIGINAL = 1	31%	55%
TRACK =0	30%	51%
TRACK =1	19%	36%
COMPLEXITY = 0	33%	51%
COMPLEXITY = 1	9%	31%
SUM = 0	38%	63%
SUM = 1	28%	47%
SUM > 1	12%	25%
SUM > 1 & COMPLEXITY = 1	4.5%	18%

Table 4 presents results analogous to those found in Table 2. The first model is again a "baseline" model containing the four control variables. As above, the coefficient for NETGT9 is positive and statistically significant (β= 0.70, z = 2.06, p < 0.05). Thus, again the data suggest that networks that produced more new series were much more likely to remove shows in the first two seasons compared to networks with fewer shows in the sample. What differs, however, is the statistical significance of the coefficient for YEAR (β= 0.16, z = 2.06, p < 0.05). This indicates that there is a trend toward higher rates of cancellation in their first two years. In Model 2, we observe that all three independent variables are statistically significant at the p <0.05 level or better (all 1-tailed). Specifically, ORIGINAL is positive and significant (p < 0.01) indicating that there is a penalty associated with original versus adapted series. In short, series that were adapted had a significantly lower chance of being canceled inside of the first two seasons. Again, this is the opposite of what we expected and what was found in the prior literature. That TRACK is negative and significant (p< 0.05) indicates that series whose creator(s) had prior success were much less likely to get canceled in the first two seasons.

Table 4. Logistic Regression where the Dependent Variable is a Categorical Variable Coded "1" if a Series was Cancelled after its 1st or 2nd Season and Coded "0" Otherwise

	Model 1	Model 2	Model 3	Model 4
YEAR	0.16*	0.20**	0.21**	0.20**
	(2.06)	(2.48)	(2.52)	(2.46)
LOGEPS	-0.98	-1.00	-0.78	-0.62
	(-0.85)	(-0.80)	(-0.64)	(-0.51)
CRIME	0.07	0.14	0.14	0.14
	(0.20)	(0.38)	(0.38)	(0.39)
NETGT9	0.70*	0.88**	0.87**	0.81**
	(2.06)	(2.45)	(2.41)	(2.26)
ORIGINAL		0.77*		
		(2.24)		
TRACK		-0.70*		
		(-1.71)		
COMPLEXITY		-1.23**		
		(-2.84)		
SUM			-0.87****	
			(-3.85)	
SUM1				-0.90*
				(-2.42)
SUM23				-1.81****
				(-3.77)
N	172	172	172	172
Chi-sq	8.8*	25.8***	25.4****	25.3****
Model df	4	7	5	6
Log Like	-114.5	-106.0	-106.2	-106.3
Pseudo-R2	3.7%	10.8%	10.7%	10.6%

Legend: # p< 0.10, 1-tailed, * p < 0.05, ** p< 0.01, *** p< 0.001 all 1-tailed

Finally, again COMPLEXITY is the most highly significant of the three variables (β= -1.23, z= -2.84, p< 0.01). This indicates that pilot episodes whose scripts were highly cognitively complex were much less likely to get cancelled in the first two seasons compared to those series whose pilots were not highly complex. Models 3 and 4 again indicate that there is an additive effect associated with the three factors—one of which, ORIGINALITY, was reversed. Specifically, in Model 3 we observe that the coefficient on SUM is negative and highly significant (β= -0.87, z = -3.85, p > 0.0001) while in Model 4, the coefficients for SUM1 and SUM23 are negative and statistically significant at the p<0.05 and p< 0.0001 levels, respectively. Thus, having more than one factor present is much better than having only factor present which is, in turn, much better for survival than having none. This fact is borne out by data in Table 3 that shows that 64%

of new dramatic series (nearly 2 in 3) with none of the three factors present were cancelled within two years whereas only 26% (better than 1 in 4) were cancelled when two or more of these factors were present. Similarly, slightly more than half of non-complex series were cancelled within two seasons whereas fewer than one-third of cognitively complex series were cancelled in the same time frame. Even more noteworthy is the fact that when both SUM23 and COMPLEXITY were equal to 1, only 20% of the series were cancelled within two seasons. That's more than a 3-fold differences in cancellation rates compared to those series lacking all three factors.

Taken together, the results of the two regression models strongly support two of our three hypotheses. Specifically, both complexity and creator track record are positively associated with series performance when that performance is measured by cancellation rates in the first and the first two seasons. However, our prediction that the originality of series would also have a positive impact on performance was not supported. Not only that, it was soundly refuted due to the fact that adapted series had the effect that we anticipated for originality—the exact opposite result. Still, we did find that these three factors combined such that new series that lacked all three factors, i.e. that were original, whose creators had no track record, and whose pilot script was not complex, were at very high risk of cancelation relative to all others. In the next section we discuss some of the implications of these results.

5. Conclusion

We began this study by noting the fact that new television series have extremely high failure rates—often as high as 80% in the first two seasons. We also noted that costs of television production are high and getting higher and that the competitive landscape is becoming crowded with new and old offerings. Taken together, it came as a surprise—to us at least—that the matter of these high failure rates and attendant costs had until very recently failed to attract much attention from academic researchers. As far as we are aware, the only research in the last two decades that directly investigated the determinants of success of new televisions series is that of Hunter and colleagues (Hunter, Smith, & Chinta, 2016; Hunter, Chinta, Smith, et al, 2016; and Hunter & Breen, 2017). Taken together with this study, a clearer picture of these determinants begins to appear. The most robust finding across all four studies is the positive and highly significant impact of cognitive complexity on a new series' performance when measured in a variety of ways. Those results hold not only for new series on the four major networks, but also for new series on over 30 cable networks and streaming services. In addition, these results parallel similar work by Hunter, Smith & Singh (2016) that found cognitive complexity positively associated with opening weekend box office revenue.

A second robust result is that associated with the track record of the creator or creative team of new series. In all four studies where this measure has been investigated, the results are positive and significant, though not as strongly as cognitive complexity. This finding also parallels the aforementioned study on opening weekend box office: that study also found a screenwriter's track record to have a positive and significant effect, but again, one weaker than that of cognitive complexity. The least robust result concerns the matter of originality of the premise of the story. In the studies on new series on network television, originality is positively associated with viewership levels in the first season, first season length and number of episodes in the first season. But in the cable sample, adapted premises out-performed. Interestingly, that's the same finding that Hunter, Smith, and Singh (2016) had in their analyses of films—adaptations outperformed original screenplays. What this last fact suggests is that originality matters but that the domain matters as well. Future research should seek to explain why films and cable television are more similar with regard to originality relative to network television series.

Finally, we should note how our results compare to the only other study in the last 25 years that directly considered determinants of success of new television shows. The study in question is Bielby & Beilby (1994) which has the title *All Hits Are Flukes: Institutionalized Decision-Making and the Rhetoric of Network Prime-Time Program Development*. In that study, the authors studied the content and consequences of claims made by network programmers concerning pilots that were being considered for upcoming fall schedules. They reported that programmer's claims linked to creator reputation increased the likelihood of new series being selected for the network's prime time schedule while claims linked to originality of series premise did not. Further, they found that neither group of claims predicted which series would succeed commercially. Those findings confirmed the conventional wisdom that "all hits are flukes" (ibid, p. 1290), that is to say, that there are no formulas for success, no single factor or combination of factors that predict commercial success of new television series. And for more than two decades, this conventional wisdom has remained unchallenged. Our findings, combined with those of related prior studies don't just challenge this conventional wisdom: they upend it. In particular, herein we have shown that new series with high level of cognitive complexity plus either a creator with a track record or an adapted premise had a 95% chance of getting a second season and an 80% chance of reaching the second season. Moreover, all eight shows in our sample that had both high cognitive complexity *and* a creator with a track record of success survived its first season. As noted earlier, these same two factors have been also shown in prior studies to positively predict a variety of commercial outcomes in new series appearing on the four major networks. Thus, it appears that even in the highly competitive context of today's television industry, hits are anything but flukes.

References

Bielby, W. T., & Bielby, D. D. (1994). "All Hits Are Flukes": Institutionalized Decision Making and the Rhetoric of Network Prime-Time Program Development. *American Journal of Sociology*, *99*(5), 1287-1313. https://doi.org/10.1086/230412

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). Ucinet for Windows: Software for social network analysis.

Danaher, P., & Dagger, T. (2012). Using a nested logit model to forecast television ratings. *International Journal of Forecasting*, *28*(3), 607-622. https://doi.org/10.1016/j.ijforecast.2012.02.008

Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2007). From story line to box office: A new approach for green-lighting movie scripts. *Management Science*, *53*(6), 881-893. https://doi.org/10.1287/mnsc.1060.0668

Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2014). Assessing box office performance using movie scripts: A kernel-based approach. *IEEE Transactions on Knowledge and Data Engineering*, *26*(11), 2639-2648. https://doi.org/10.1109/TKDE.2014.2306681

Hunter, S. (2014). A novel method of network text analysis. *Open Journal of Modern Linguistics*, *4*(02), 350-366. https://doi.org/10.4236/ojml.2014.42028

Hunter, S., & Breen (2017). W(h)ither the Full Season: An Empirical Model for Predicting Duration of New Television Series' First Season. Advances in Journalism and Communication, in press

Hunter, S., Chinta, R., Smith, S., Shamim, A., & Bawazir, A. (2016). Moneyball for TV: A Model for Forecasting the Audience of New Dramatic Television Series. *Studies in Media and Communication*, *4*(2), 13-22. https://doi.org/10.11114/smc.v4i2.1611

Hunter, S., Smith, S., & Chinta, R. (2016). Predicting New TV Series Ratings from their Pilot Episode Scripts. *International Journal of English Linguistics*, *6*(5), 1-11. https://doi.org/10.5539/ijel.v6n5p1

Hunter, S., Smith, S., & Singh, S. (2016). Predicting box office from the screenplay: A text analytical approach. *Journal of Screenwriting*, *7*(2), 135-154. https://doi.org/10.1386/josc.7.2.135_1

Littleton, C. (2015, August 7). FX Networks Chief John Landgraf: 'There Is Simply Too Much Television'. *Variety*. Retrieved April 10, 2017, from http://variety.com/2015/tv/news/tca-fx-networks-john-landgraf-wall-street-1201559191/

Nathanson, J. (2013, October 17). The Economics of a Hit Show. *Priceonomics*. Retrieved April 6, 2017 from http://priceonomics.com/the-economics-of-a-hit-tv-show/

Porter, R. (2016, June 1). Failure factor: How bad was the 2015-16 TV season for new shows? Retrieved April 6, 2017, from http://tvbythenumbers.zap2it.com/more-tv-news/failure-factor-how-bad-was-the-2015-16-tv-season-for-new-show/

Ryan, M. (2016, December 21). TV Peaks Again in 2016: Could It Hit 500 Shows in 2017? Variety. Retrieved March 30, 2017, from http://variety.com/2016/tv/news/peak-tv-2016-scripted-tv-programs-1201944237/

Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry*, 405-414. https://doi.org/10.2307/2785779

Stelter, B. (2012, May 13). Few TV Shows Survive a Ruthless Proving Ground. *New York Times*. Retrieved from http://www.nytimes.com/2012/05/14/business/media/few-tv-shows-survive-a-ruthless-proving-ground.html

Tinsley, H. E., Kass, R. A., Moreland, J. R., & Harren, V. A. (1983). A longitudinal study of female college students' occupational decision making. *The Career Development Quarterly*, *32*(2), 89-102. https://doi.org/10.1002/j.2164-585x.1983.tb01563.x

Watkins, C. (2011). The American Heritage Dictionary of Indo-European Roots. 2011.

Copyrights