

# Reinterpretation of LHC results

## Current experiences, and lessons for EFT

**Andy Buckley, University of Glasgow**

for the LHC BSM Reinterpretation Forum, et al

**1st LHC EFT WG general meeting**  
**20 October 2020**



University  
of Glasgow



THE ROYAL  
SOCIETY

# Interpretation and re-interpretation

- ❖ **What's the difference? Could slice many ways, but:**
  - **Interpretation** ~ mapping measurements to model parameters in-experiment
  - **Reinterpretation** ~ enabling and performing *post hoc* interpretations externally, from public information
- ❖ **Typical consequences:**
  - Interpretation by experimentalists, using experiment toolchains: focus on most impactful BSM inferences possible from simpler/simplified models
  - Experiment-based analysis combinations can be detailed in systematics, but are limited in sampling sophistication or numbers of sim+reco events possible within CPU budget
  - Reinterpretations increasingly focus on more general models, param spaces where adaptive sampling is essential... but lack the control over experiment effects. Unfolding helps (?)



# Reinterpretation Forum report

## ❖ LHC BSM Reinterpretation Forum updated status report + recommendations: [arXiv:2003.07868](https://arxiv.org/abs/2003.07868)

- Reflections on data & metadata, available tools, and concrete recommendations for expt and pheno
- Covers direct-search reinterpretation (most of current pheno effort), plus GPD and flavour measurements
- Fundamentally same limiting issues in direct & indirect (after all, search SRs  $\sim d\sigma/dX$  bins):
  - background estimates and correlation structures needed for composite likelihoods;
  - parameter-space sampling challenges.

## ❖ Addressing HEP's version of the "reproducibility crisis": *we're better, but not exempt!*

CERN-LPCC-2020-001, FERMILAB-FN-1098-CMS-T, Imperial/HEP/2020/RIF/01

### Reinterpretation of LHC Results for New Physics: Status and Recommendations after Run 2

We report on the status of efforts to improve the reinterpretation of searches and measurements at the LHC in terms of models for new physics, in the context of the LHC Reinterpretation Forum. We detail current experimental offerings in direct searches for new particles, measurements, technical implementations and Open Data, and provide a set of recommendations for further improving the presentation of LHC results in order to better enable reinterpretation in the future. We also provide a brief description of existing software reinterpretation frameworks and recent global analyses of new physics that make use of the current data.

Waleed Abdallah,<sup>1,2</sup> Shehu AbdusSalam,<sup>3</sup> Azar Ahmadov,<sup>4</sup> Amine Ahriche,<sup>5,6</sup> Gaël Alguero,<sup>7</sup> Benjamin C. Allanach,<sup>8,\*</sup> Jack Y. Araz,<sup>9</sup> Alexandre Arbey,<sup>10,11</sup> Chiara Arina,<sup>12</sup> Peter Athron,<sup>13</sup> Emanuele Bagnaschi,<sup>14</sup> Yang Bai,<sup>15</sup> Michael J. Baker,<sup>16</sup> Csaba Balazs,<sup>13</sup> Daniele Barducci,<sup>17,18</sup> Philip Bechtle,<sup>19,\*</sup> Aoife Bharucha,<sup>20</sup> Andy Buckley,<sup>21,\*</sup> Jonathan Butterworth,<sup>22,\*</sup> Haiying Cai,<sup>23</sup> Claudio Campagnari,<sup>24</sup> Cari Cesarotti,<sup>25</sup> Marcin Chrzaszcz,<sup>26</sup> Andrea Coccaro,<sup>27</sup> Eric Conte,<sup>28,29</sup> Jonathan M. Cornell,<sup>30</sup> Louie D. Corpe,<sup>22</sup> Matthias Danninger,<sup>31</sup> Luc Darmé,<sup>32</sup> Aldo Deandrea,<sup>10</sup> Nishita Desai,<sup>33,\*</sup> Barry Dillon,<sup>34</sup> Caterina Doglioni,<sup>35</sup> Matthew J. Dolan,<sup>16</sup> Juhi Dutta,<sup>1,36</sup> John R. Ellis,<sup>37</sup> Sebastian Ellis,<sup>38</sup> Farida Fassi,<sup>39</sup> Matthew Feickert,<sup>40</sup> Nicolas Fernandez,<sup>40</sup> Sylvain Fichet,<sup>41</sup> Thomas Flacke,<sup>42</sup> Benjamin Fuks,<sup>43,44,\*</sup> Achim Geiser,<sup>45</sup> Marie-Hélène Genest,<sup>7</sup> Akshay Ghalsasi,<sup>46</sup> Tomas Gonzalo,<sup>13</sup> Mark Goodsell,<sup>43</sup> Stefania Gori,<sup>46</sup> Philippe Gras,<sup>47</sup> Admir Greljo,<sup>11</sup> Diego Guadagnoli,<sup>48</sup> Sven Heinemeyer,<sup>49,50,51</sup> Lukas A. Heinrich,<sup>11,\*</sup> Jan Heisig,<sup>12,\*</sup> Deog Ki Hong,<sup>52</sup> Tetiana Hryn'ova,<sup>53</sup> Katri Huitu,<sup>54</sup> Philip Ilten,<sup>55</sup> Ahmed Ismail,<sup>56</sup> Adil Jueid,<sup>57</sup> Felix Kahlhoefer,<sup>58,\*</sup> Jan Kalinowski,<sup>59</sup> Jernej F. Kamenik,<sup>60,34</sup> Deepak Kar,<sup>61</sup> Yevgeny Kats,<sup>62</sup> Charanjit K. Khosa,<sup>63</sup> Valeri Khoze,<sup>64</sup> Tobias Klöckl,<sup>19</sup> Damir Kozlov,<sup>65</sup> Kyoungchul Kong,<sup>66</sup> Wojciech Kotlarski,<sup>67</sup> Michael Krämer,<sup>58</sup>

arXiv:2003.07868v3 [hep-ph] 21 Jul 2020

# RiF headline recommendations

## ❖ For everyone:

- **Broken-down SM (background) yield estimates with full correlations (in standard format, via HepData) and generation/fit info.** Alternatively/complementarily full likelihoods via **pyhf JSON**

## ❖ For searches:

- Publish **cut-flows for multiple benchmark models and all signal regions, with full generation info**
- Publish (analysis-specific) **detector biasing/smearing functions + digitized trigger/reco efficiencies**
- **Analysis preservation** as (at least) public pseudocode, **ideally publicly executable**
- + specifics for LLPs and simplified models

## ❖ For measurements:

- **Define fiducial cross-sections in terms of final-state particles: don't subtract irreducible backgrounds!**
- Make explicit any **limitations on the validity of unfolded fiducial cross-section definitions**
- **A Rivet routine should be provided**
- Document **highest observed points on tails of distributions**
- **Visible cross-sections for STXS categories should be published**
- + specifics for flavour and EFT fits

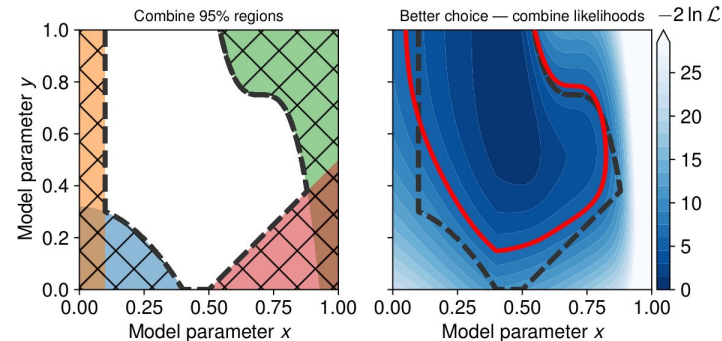
# Composite likelihoods & computational tools

## ❖ Composite likelihoods are the only way forward

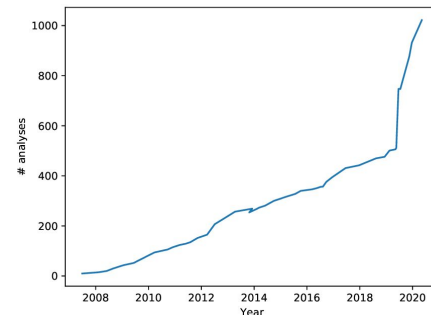
- Exclusion overlays misrepresent, and restrict to specific, few-param models: need to combine analyses coherently

## ❖ Bustling marketplace of tools for all aspects

- **Generation:** UFO + MG5/Sherpa/Herwig + SHGs
- **Detector emulation:** Delphes/Rivet/MA5/ColliderBit fast-sims
- **Analysis:** MadAnalysis5, ColliderBit, Rivet, CheckMATE (different current emphases re. search, measurement, LLPs)
- **Likelihood evaluation:** above, plus Contur, SModelS
- **All the above, *a la* experiment:** RECAST
- **(Global) fits:** Gambit, MasterCode, HEPfit, SMEFIT, TopFitter, EFTfitter, ...



GAMBIT et al, in prep

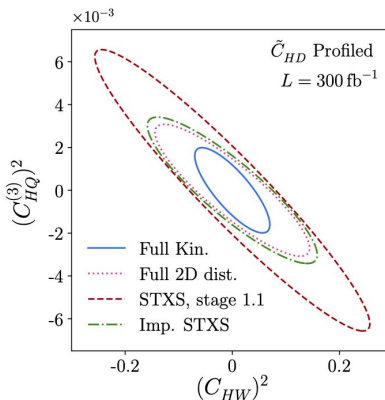


Rivet analyses vs time

# Targetted or generic observables?

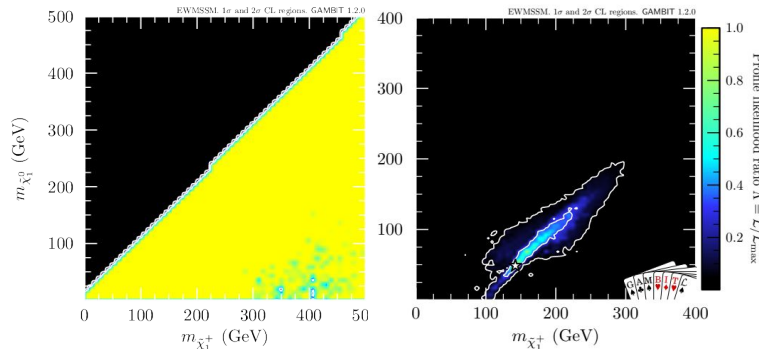
## ❖ Better to target specific models/operators, or interpret from a model-independent set?

- Variables defined at ME level can maximise theoretical power, often using ML methods
- Needs full-sim, reweighting approaches between template set (e.g. EFT score reweighting), “bakes in” EFT order and MC: model-dependence & fragility
- Differential cross-sections generic and fast, but less powerful in individual constraining power



## ❖ False dichotomy: we need both!

- Each makes assumptions: 2 versions  $\Rightarrow$  comparison
- Overall interpretation needs to be “full” model spaces... but those will evolve with time & tech
- Gambit example: EWMSSM mostly-excluded simplified model  $\rightarrow$  no constraint with slight relaxation. Convenient 1D/2D slices can mislead



# Backgrounds

## ❖ Reporting background estimates

- Background-modelling reporting is key: many samples are prohibitive in person and/or CPU time
- Increased RiF emphasis on reporting of background estimates (incl in measurements) and generation setups
- E.g. run cards or even (LHE) event-sample sharing

## ❖ And correlations in those estimates

- See next slide!

## ❖ But don't subtract them!

- Subtraction is a potentially fatal source of SM bias
- Similar assumptions in STXS. Assumptions built-in, vs fiducial where assumptions can be delayed or avoided. Strong case for also publishing fiducial-style counts

Bin	$H_T^{\text{min}}$ [GeV]	$H_T$ [GeV]	$N_{\text{pt}}$	$N_{\text{jet}}$	Loss- $\epsilon/\mu$	$\tau \rightarrow \text{had}$	$Z \rightarrow \nu\bar{\nu}$	QCD	Total Pred.	Obs.
81	300-350	300-500	7-8	0	1.88 <sup>+0.08</sup> <sub>-0.04</sub>	0.97 <sup>+0.65</sup> <sub>-0.31</sub>	3.79 <sup>+1.20</sup> <sub>-1.19</sub>	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	6.59 <sup>+1.91</sup> <sub>-1.34</sub>	8
82	300-350	500-1000	7-8	0	14.11 <sup>+1.44</sup> <sub>-1.22</sub>	18.61 <sup>+1.81</sup> <sub>-1.93</sub>	28.47 <sup>+2.97</sup> <sub>-3.38</sub>	6.23 <sup>+1.12</sup> <sub>-0.61</sub>	67.72 <sup>+3.67</sup> <sub>-4.85</sub>	69
83	300-350	1000-	7-8	0	10.01 <sup>+2.26</sup> <sub>-1.63</sub>	8.30 <sup>+1.37</sup> <sub>-1.31</sub>	7.89 <sup>+1.88</sup> <sub>-1.96</sub>	10.06 <sup>+2.39</sup> <sub>-2.22</sub>	36.26 <sup>+3.85</sup> <sub>-3.78</sub>	27
84	350-500	350-500	7-8	0	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	2.33 <sup>+1.17</sup> <sub>-1.11</sub>	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	2.39 <sup>+1.13</sup> <sub>-1.11</sub>	3
85	350-500	500-1000	7-8	0	15.61 <sup>+1.21</sup> <sub>-1.24</sub>	17.91 <sup>+1.23</sup> <sub>-1.28</sub>	38.57 <sup>+1.63</sup> <sub>-1.88</sub>	3.69 <sup>+1.52</sup> <sub>-1.75</sub>	75.77 <sup>+3.98</sup> <sub>-4.48</sub>	78
86	350-500	1000-	7-8	0	10.09 <sup>+1.97</sup> <sub>-1.85</sub>	9.15 <sup>+1.63</sup> <sub>-1.57</sub>	14.64 <sup>+2.19</sup> <sub>-1.87</sub>	5.60 <sup>+1.47</sup> <sub>-1.35</sub>	39.49 <sup>+4.29</sup> <sub>-4.05</sub>	35
87	500-750	500-1000	7-8	0	2.68 <sup>+1.10</sup> <sub>-0.95</sub>	1.11 <sup>+0.61</sup> <sub>-0.40</sub>	6.78 <sup>+1.56</sup> <sub>-1.07</sub>	0.32 <sup>+0.48</sup> <sub>-0.30</sub>	10.90 <sup>+2.56</sup> <sub>-1.93</sub>	14
88	500-750	1000-	7-8	0	2.25 <sup>+0.66</sup> <sub>-0.40</sub>	2.77 <sup>+0.86</sup> <sub>-0.72</sub>	10.29 <sup>+2.06</sup> <sub>-1.92</sub>	0.70 <sup>+0.19</sup> <sub>-0.18</sub>	16.01 <sup>+2.63</sup> <sub>-2.48</sub>	22
89	750-	750-1500	7-8	0	0.27 <sup>+0.41</sup> <sub>-0.11</sub>	0.03 <sup>+0.46</sup> <sub>-0.01</sub>	0.88 <sup>+0.62</sup> <sub>-0.34</sub>	0.04 <sup>+0.07</sup> <sub>-0.04</sub>	1.22 <sup>+1.15</sup> <sub>-0.34</sub>	3
90	750-	1500-	7-8	0	0.20 <sup>+0.47</sup> <sub>-0.10</sub>	0.43 <sup>+0.57</sup> <sub>-0.20</sub>	0.40 <sup>+0.40</sup> <sub>-0.11</sub>	0.13 <sup>+0.12</sup> <sub>-0.12</sub>	1.16 <sup>+1.12</sup> <sub>-0.33</sub>	2
91	300-350	300-500	7-8	1	0.65 <sup>+0.78</sup> <sub>-0.33</sub>	1.43 <sup>+0.69</sup> <sub>-0.48</sub>	1.76 <sup>+0.60</sup> <sub>-0.38</sub>	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	3.84 <sup>+3.56</sup> <sub>-1.88</sub>	5
92	300-350	500-1000	7-8	1	17.45 <sup>+2.23</sup> <sub>-2.20</sub>	25.39 <sup>+2.16</sup> <sub>-2.12</sub>	13.30 <sup>+2.19</sup> <sub>-2.40</sub>	1.49 <sup>+0.55</sup> <sub>-0.42</sub>	57.63 <sup>+4.96</sup> <sub>-4.71</sub>	55
93	300-350	1000-	7-8	1	4.92 <sup>+1.30</sup> <sub>-1.04</sub>	5.70 <sup>+1.15</sup> <sub>-1.05</sub>	3.70 <sup>+0.90</sup> <sub>-0.82</sub>	6.08 <sup>+0.63</sup> <sub>-0.57</sub>	20.40 <sup>+2.69</sup> <sub>-2.49</sub>	26
94	350-500	350-500	7-8	1	0.63 <sup>+1.07</sup> <sub>-0.30</sub>	0.01 <sup>+0.46</sup> <sub>-0.00</sub>	1.09 <sup>+0.56</sup> <sub>-0.33</sub>	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	1.73 <sup>+1.61</sup> <sub>-0.33</sub>	8
95	350-500	500-1000	7-8	1	21.48 <sup>+1.93</sup> <sub>-1.92</sub>	21.94 <sup>+1.74</sup> <sub>-1.82</sub>	17.95 <sup>+1.81</sup> <sub>-1.84</sub>	3.93 <sup>+1.67</sup> <sub>-1.46</sub>	65.31 <sup>+3.60</sup> <sub>-4.25</sub>	42
96	350-500	1000-	7-8	1	7.13 <sup>+1.41</sup> <sub>-1.07</sub>	7.88 <sup>+1.38</sup> <sub>-1.04</sub>	6.89 <sup>+1.29</sup> <sub>-1.31</sub>	2.80 <sup>+1.16</sup> <sub>-1.06</sub>	24.70 <sup>+3.18</sup> <sub>-2.66</sub>	14
97	500-750	500-1000	7-8	1	1.28 <sup>+0.82</sup> <sub>-0.67</sub>	2.23 <sup>+0.83</sup> <sub>-0.70</sub>	3.13 <sup>+0.82</sup> <sub>-0.71</sub>	0.09 <sup>+0.39</sup> <sub>-0.30</sub>	6.73 <sup>+1.89</sup> <sub>-1.60</sub>	9
98	500-750	1000-	7-8	1	7.00 <sup>+2.91</sup> <sub>-1.13</sub>	3.98 <sup>+1.23</sup> <sub>-0.84</sub>	4.89 <sup>+1.61</sup> <sub>-1.08</sub>	0.42 <sup>+0.33</sup> <sub>-0.32</sub>	16.29 <sup>+4.39</sup> <sub>-3.67</sub>	10
99	750-	750-1500	7-8	1	0.43 <sup>+0.32</sup> <sub>-0.16</sub>	0.40 <sup>+0.57</sup> <sub>-0.23</sub>	0.40 <sup>+0.29</sup> <sub>-0.19</sub>	0.17 <sup>+0.43</sup> <sub>-0.20</sub>	1.40 <sup>+1.21</sup> <sub>-0.37</sub>	1
100	750-	1500-	7-8	1	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	0.02 <sup>+0.46</sup> <sub>-0.01</sub>	0.21 <sup>+0.21</sup> <sub>-0.06</sub>	0.02 <sup>+0.04</sup> <sub>-0.02</sub>	0.25 <sup>+1.01</sup> <sub>-0.06</sub>	2
101	300-350	300-500	7-8	2	0.27 <sup>+0.06</sup> <sub>-0.03</sub>	0.33 <sup>+0.30</sup> <sub>-0.08</sub>	0.69 <sup>+0.20</sup> <sub>-0.26</sub>	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	1.19 <sup>+1.20</sup> <sub>-0.30</sub>	2
102	300-350	500-1000	7-8	2	19.23 <sup>+4.30</sup> <sub>-2.72</sub>	20.14 <sup>+3.99</sup> <sub>-3.12</sub>	4.62 <sup>+1.32</sup> <sub>-1.23</sub>	0.38 <sup>+0.36</sup> <sub>-0.44</sub>	44.37 <sup>+6.63</sup> <sub>-5.74</sub>	28
103	300-350	1000-	7-8	2	1.72 <sup>+0.94</sup> <sub>-0.33</sub>	3.76 <sup>+1.94</sup> <sub>-1.48</sub>	1.30 <sup>+0.44</sup> <sub>-0.44</sub>	2.37 <sup>+0.41</sup> <sub>-0.35</sub>	9.15 <sup>+3.91</sup> <sub>-1.57</sub>	13
104	350-500	350-500	7-8	2	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	0.01 <sup>+0.46</sup> <sub>-0.01</sub>	0.38 <sup>+0.22</sup> <sub>-0.17</sub>	0.00 <sup>+0.00</sup> <sub>-0.00</sub>	0.39 <sup>+0.22</sup> <sub>-0.17</sub>	1

No-one should ever have to retype this again...



# Correlations

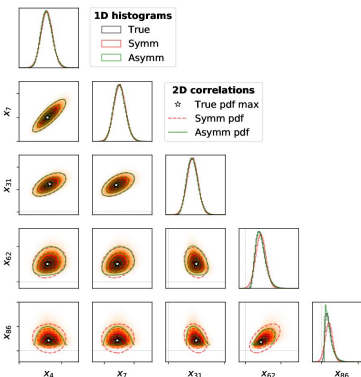
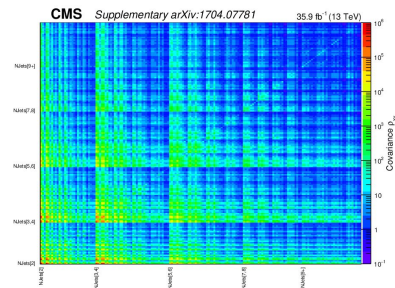
## ❖ Correlations are critical to combination

- From: **modelling systematics, experiment systematics, statistics, event-sharing**
- **Without correlation knowledge, can only safely use one bin from each “analysis pool”: a waste of composite power**
- Also important to stabilise fits: *which* single bin to use can be unstable, leading to “flip-flopping”

## ❖ Reporting

- Simplest forms: **error-sources** (Gaussian) or **Simplified Likelihood** (Gaussian on Poisson rate)
- Breaks down, particularly in low-rate SRs: **NSL** adds skew via  $N_{\text{bin}}$  additional numbers. **pyhf** now offers a compact format for full templated-likelihood reporting
- Machinery all exists, but publishing is far from universal: **still many analyses publish data without correlations (and often/usually much later than the paper)**

RE	P P → JETS						
COS PHI	TEEC						
-1 - -0.96	10.5165 ±0.00779481	stat	+0.0117651 -0.0113337	sys_jesNp1	+0.0034308 -0.00335944	sys_jesNp2 + 71 more errors	<a href="#">Show all</a>
-0.96 - -0.92	0.716955 ±0.00468718	stat	+0.00357006 -0.00426249	sys_jesNp1	+0.00165822 -0.00195598	sys_jesNp2 + 71 more errors	<a href="#">Show all</a>
-0.92 - -0.88	0.322052 ±0.00299636	stat	+0.00184137 -0.00189796	sys_jesNp1	+0.000834941 -0.00104189	sys_jesNp2 + 71 more errors	<a href="#">Show all</a>

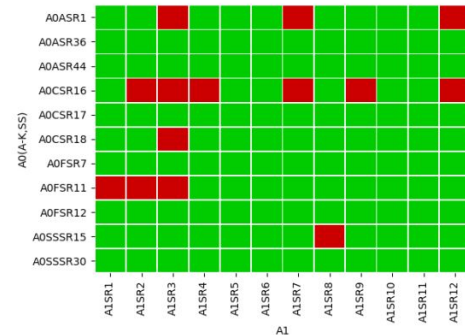




# Correlations between analyses/experiments

## ❖ Cross-analysis and cross-experiment systematics

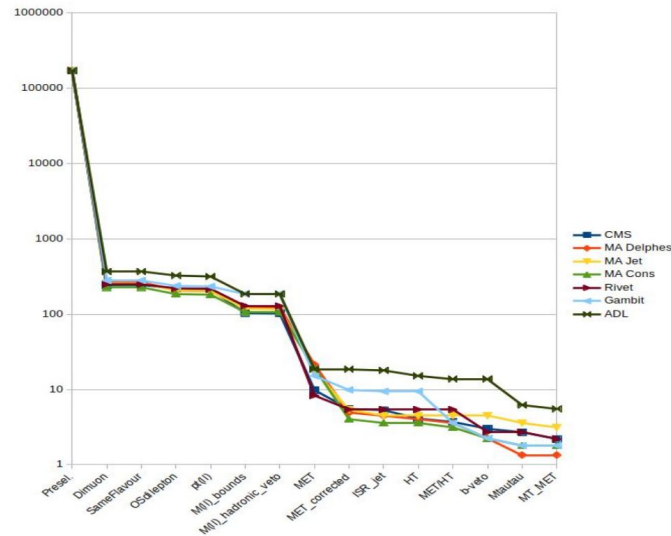
- These will enter at some point... how reliable/optimal are guesses of orthogonality? How much power is missed?
- **Not easy:**
  - calibrations & systematics evolve (with same name)
  - phase-spaces may have different best-fit nuisances
  - same methods/MC underpinning different expts
- **Even if identified, feasibility of uber-fits?**
- Not too early: the era of rapidly superseded direct searches is over (and has never been the case for measurements)
- Les Houches 2019: <https://arxiv.org/abs/2002.12220>  
**Top-down & empirical estimates of event-sharing;** some formal overlaps are ok.



# Analysis preservation

## ❖ Full-fat or diet recasting?

- RECAST vs fastcast paradigms: *not* mutually exclusive
  - Full detail needed for full reproducibility
  - Unfolding/smearing → speed → adaptive scanning
  - Staging: use fast interpretations in high dimensionality to point at areas of interest for forensic reinterpretation
- Speed challenging for MC gen: low-acceptance analyses require *many* events, and high-order QCD is CPU-punitive
- Reweighting techniques interesting as an alternative, esp in EFT where dependences are well-behaved/factorisable
- Big ATLAS effort in “forensic” preservation of native analysis code; longer-term fast-analysis preservation needs same formality and timeliness. Doesn't come for free...
- Also: independent implementations find analysis bugs!
- Preserving MVAs — technical robustness, physical interpretability at particle/fastsim level?



### Rivet analysis coverage (no searches, no heavy ion)

Rivet analyses exist for 839/4241 papers = 20%. 157 priority analyses required.

Total number of Inspire papers scanned = 7280, at 2020-06-26

Breakdown by identified experiment (in development):

Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	$e^+e^- (\geq 12 \text{ GeV})$
Rivet wanted (total):	72	113	130	183	43	461	765
Rivet REALLY wanted:	17	43	64	9	0	13	1
Rivet provided:	14/86 = 16%	133/246 = 54%	73/203 = 36%	13/196 = 7%	8/51 = 16%	9/470 = 2%	166/931 = 18%

# Summary

## ❖ **Big improvements in depth of published data**

- *“The best is the enemy of the good.”* — Voltaire
- Recent years have provided lots of well-judged “good” ⇒ lots of reinterpretation
- Entering a new, “statistically stable” experimental era: let’s learn from experience
- RECAST vs fastcast paradigms: not mutually exclusive, can/should stage interpretations to roughly point at areas of interest for forensic reinterpretation

## ❖ **Analysis and likelihood/correlation preservation too**

- Correlation publication becoming standard practice in measurements
- Particularly big movements re. pyhf and RECAST in ATLAS searches
- Need similar effort in experiment: pheno reimplementations can’t keep up
- Perceived tension (?) between full and fast approaches, cf. targetted and generic observables. We probably need both... more work, but not 2x

## ❖ **Many tools for all stages of (re)interpretation**

## ❖ **Duplication and format mismatches cost person-power:**

- expt and pheno need to coordinate to maximise measurement power *and* longevity